

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 00:24:35 ; Search time 162 Seconds
(without alignments)
229.191 Million cell updates/sec

Title: US-10-632-414-3
Sequence: 1 ATTTSTGSAFVNQHLGSH.....IVEQCTCSISLVLENYCG 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.GeneSeq.16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	514	100.0	96	ADN34535	Adn34535 Preproins
2	508	98.8	96	ADN34534	Adn34534 Preproins
3	504	98.1	96	AAV08005	AAV08005 Human pro
4	504	98.1	96	AAO17831	AAO17831 Human pro
5	504	98.1	97	AAV78661	AAV78661 Proinsulin
6	501	97.5	97	AAV15472	AAV15472 NcoI-Sali
7	498	96.9	96	AAV08004	AAV08004 Human pro
8	498	96.9	96	AAO17830	AAO17830 Human pro
9	498	96.9	97	AAV68898	AAV68898 Human pro
10	498	96.9	97	AAV93414	AAV93414 Human pro
11	497.5	96.8	98	AAV51222	AAV51222 Human pre
12	497	96.7	99	AAV51223	AAV51223 Human pre
13	496.5	96.6	100	AAV51224	AAV51224 Human pre
14	496	96.5	97	AAV93416	AAV93416 Human ins
15	492	95.7	97	AAV93419	AAV93419 Human ins
16	492	95.7	97	AAV93415	AAV93415 Human ins
17	488	89.1	153	AAV53589	AAV53589 Human pre
18	455	88.5	92	AAV20467	AAV20467 Yeast alp
19	455	88.5	93	AAV90102	AAV90102 Synthetic
20	455	88.5	96	ADL24437	ADL24437 Modified
21	454	88.3	110	AAV10042	AAV10042 Sequence
22	454	88.3	110	AAV10053	AAV10053 Sequence
23	454	88.3	110	AAV40309	AAV40309 Sequence
24	454	88.3	110	AAV06608	AAV06608 Human pre
25	454	88.3	110	ADV77582	ADV77582 Human ins

26	454	88.3	110	3	AAV44367	AAV44367 Human pro
27	454	88.3	110	3	AAV70366	AAV70366 Human ins
28	454	88.3	110	3	AAV26765	AAV26765 Human pre
29	454	88.3	110	3	AAV06144	AAV06144 Human ins
30	454	88.3	110	4	AAV10337	AAV10337 Human pre
31	454	88.3	110	4	AAV35424	AAV35424 Secretary
32	454	88.3	110	5	AAV65677	AAV65677 Human pro
33	454	88.3	110	5	AAV60634	AAV60634 Human ins
34	454	88.3	110	5	ABG31590	ABG31590 Human pre
35	454	88.3	110	6	ABV55862	ABV55862 Human ins
36	454	88.3	110	6	ADA09218	ADA09218 Human pre
37	454	88.3	110	7	ADV56710	ADV56710 Human pro
38	454	88.3	110	7	ADV46938	ADV46938 Human pro
39	454	88.3	110	7	ADV57650	ADV57650 Human pro
40	454	88.3	110	8	ADL25930	ADL25930 Human pre
41	454	88.3	110	8	ADL24433	ADL24433 Human ins
42	454	88.3	110	8	ADN00699	ADN00699 Human ins
43	454	88.3	110	8	ADL90090	ADL90090 Human ins
44	454	88.3	110	8	ADN49716	ADN49716 Human ins
45	454	88.3	110	8	ADQ14352	ADQ14352 Human pre

ALIGNMENTS

RESULT 1
ADN34535 standard; peptide; 96 AA.
ID ADN34535;
AC ADN34535;
XX 17-JUN-2004 (first entry)
DT Preproinsulin peptide fragment #3.
DE Preproinsulin; insulin; diabetes mellitus; fermentation;
XX preproinsulin; insulin; diabetes mellitus; fermentation;
KM genetically modified microorganism.
KW
XX Homo sapiens.
OS
DE DE10235168-A1.
PN 12-FEB-2004.
PD
XX 01-AUG-2002; 2002DE-01035168.
PF
XX 01-AUG-2002; 2002DE-01035168.
PR
XX (AVET) AVENTIS PHARMA DEUT GMBH.
PA Thuro H, Blumenstock H, Havenith C;
XX WPI; 2004-228442/22.
PT Purification of preproinsulin, useful for conversion to insulin,
PT comprises removing high molecular weight impurities by anion-exchange
PT then cation-exchange chromatography.
PS Claim 3; SEQ ID NO 3; 19pp; German.
XX This invention describes a novel method for purification of preproinsulin
XX which comprises removing high molecular weight substances from an aqueous
XX solution by chromatography on an anion exchanger in flow-through mode and
XX separation by chromatography on a cation exchanger in adsorption mode. A
XX method is also described for preparation of insulin by expression of
XX unfolded preproinsulin. The method is used as part of a process for
XX preparing insulin (for treating diabetes mellitus) by fermentation of
XX genetically modified microorganisms that express unfolded preproinsulin.
XX The method provides efficient removal of contaminants (particularly
XX polymeric preproinsulin) that adversely affect stability of insulin
XX during subsequent processing steps.
SQ Sequence 96 AA;

Query Match 100.0%; Score 514; DB 8; Length 96;
 Best Local Similarity 100.0%; Pred. No. 9e-49;
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60
 |||||
 DB 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60
 |||||
 QY 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYCG 96
 |||||
 DB 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYCG 96
 |||||

RESULT 2
 ADN34534
 ID ADN34534 standard; peptide; 96 AA.

AC ADN34534;

DT 17-JUN-2004 (first entry)

DE Preproinsulin peptide fragment #2.

XX preproinsulin; insulin; diabetes mellitus; fermentation;

KW genetically modified microorganism.

XX Homo sapiens.

XX DE10235168-A1.

XX 12-FEB-2004.

XX 01-AUG-2002; 2002DE-01035168.

XX 01-AUG-2002; 2002DE-01035168.

XX (AVET) AVENTIS PHARMA DEUT GMBH.

XX Thurorow H, Blumenstock H, Havenith C;

XX WPI; 2004-228442/22.

XX Purification of preproinsulin, useful for conversion to insulin,

PT comprises removing high molecular weight impurities by anion-exchange

PT then cation-exchange chromatography.

XX Claim 2; SEQ ID NO 2; 19pp; German.

XX This invention describes a novel method for purification of preproinsulin
 CC which comprises removing high molecular weight substances from an aqueous
 CC solution by chromatography on an anion exchanger in flow-through mode and
 CC separation by chromatography on a cation exchanger in adsorption mode. A
 CC method is also described for preparation of insulin by expression of
 CC unfolded preproinsulin. The method is used as part of a process for
 CC preparing insulin (for treating diabetes mellitus) by fermentation of
 CC genetically modified microorganisms that express unfolded preproinsulin.
 CC The method provides efficient removal of contaminants (particularly
 CC polymeric preproinsulin) that adversely affect stability of insulin
 CC during subsequent processing steps.

XX Sequence 96 AA;

Query Match 98.8%; Score 508; DB 8; Length 96;

Best Local Similarity 100.0%; Pred. No. 4.1e-48;
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60
 |||||
 DB 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60
 |||||
 QY 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYCG 95
 |||||

DB 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYCG 95

RESULT 3
 AAY08005
 ID AAY08005 standard; peptide; 96 AA.

XX AAY08005;

XX 20-MAR-2003 (revised)

XX 08-JUL-1999 (first entry)

XX Human proinsulin protein fragment 2.

XX Insulin; human; A chain; B chain; pro-insulin; chaotropic agent.

XX Homo sapiens.

XX EP906918-A2.

XX 07-APR-1999.

XX 11-AUG-1998; 98EP-00115048.

XX 18-AUG-1997; 97DE-01035711.

XX (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.

XX Rubroeder FD, Keller R;

XX WPI; 1999-135401/22.

XX Preparation of insulin precursors with cystine bridges in the presence of

XX cysteine and chaotropic auxiliary.

XX Example 1; Page 12; 16pp; German.

XX This invention describes a method for the production of insulin or

CC insulin derivative precursors in an aqueous suspension which are diluted

CC with cysteine or cysteine hydrochloride to pH 11.5 and then at 55 deg. C

CC a chaotropic additive is added. This sequence represents human pro-

CC insulin which is used in the description of the invention. This patent is

CC an equivalent to HU9801886. (Updated on 20-MAR-2003 to correct DR field.)

XX Sequence 96 AA;

Query Match 98.1%; Score 504; DB 2; Length 96;

Best Local Similarity 99.0%; Pred. No. 1.1e-47;
 Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60
 |||||
 DB 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60
 |||||

QY 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYCG 96
 |||||

DB 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYCG 96
 |||||

RESULT 4
 AAO17831
 ID AAO17831 standard; peptide; 96 AA.

XX AAO17831;

XX 13-SEP-2002 (first entry)

XX Human proinsulin #2.

XX Human; insulin; cysteine bridge; kaotropic additive.

XX Homo sapiens.

PN EP0980874-A1.
 XX 23-FEB-2000.
 XX
 XX
 PF 11-AUG-1998; 99EP-00115386.
 XX
 PR 18-AUG-1997; 97DE-01035711.
 PR 11-AUG-1998; 98EP-00115048.
 XX
 PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.
 XX
 PI Rubroeder F, Keller R;
 XX WPI; 1999-135401/12.
 DR
 XX
 PT Preparation of insulin precursors with cysteine bridges in the presence of
 PT cysteine and chaotropic auxiliary.
 XX
 PS Example 2; Page 8; 20pp; German.
 XX
 CC The present invention relates to insulin or insulin derivative precursors
 CC in an aqueous suspension, which are diluted with cysteine or cysteine
 CC hydrochloride to pH 11.5 and then at 55psioc a chaotropic additive is
 CC added. The present sequence is one version of human proinsulin
 XX
 SO Sequence 96 AA;
 Query Match 98.1%; Score 504; DB 2; Length 96;
 Best Local Similarity 99.0%; Pred. No. 1.1e-47;
 Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 ATGTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADPOVQVELGGGPGA 60
 Db 1 ATGTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADPOVQVELGGGPGA 60
 Oy 61 GSLOPLALGSLQKRGIVEQCCTSIQSLYOLENYCG 96
 Db 61 GSLOPLALGSLQKRGIVEQCCTSIQSLYOLENYCG 96
 RESULT 5
 AAR78661
 ID AAR78661 standard; protein; 97 AA.
 XX
 AC AAR78661;
 XX
 DT 03-APR-1996 (first entry)
 XX
 DE Proinsulin sequence 1.
 XX
 KW Proinsulin; post-translational modification; recombinant production;
 KM protein folding; conformation.
 XX
 OS Synthetic.
 XX
 FH Key location/Qualifiers
 FT Region 1..11
 FT /label= R1
 FT /note= "a peptide of 11 amino acids"
 FT Peptide 12..41
 FT /label= R1-(B2-B29)-Y
 FT /note= "human insulin B-chain"
 FT Peptide 42..76
 FT /label= X
 FT /note= "C-peptide of human insulin"
 FT Peptide 77..97
 FT /label= Gly-R3
 FT /note= "human insulin A-chain"
 XX
 XX EP668292-A2.
 XX 23-AUG-1995.
 XX

PF 09-FEB-1995; 95EP-00101748.
 XX
 PR 18-FEB-1994; 94DE-04405179.
 XX
 PA (FARH) HOECHST AG.
 XX
 PI Obermeier R, Gerl M, Ludwig J, Sabel W;
 XX WPI; 1995-264754/38.
 DR
 XX
 PT Isolation of insulin that is correctly post-translationally processed -
 PT by reacting pro:insulin with a mercaptan in the presence of a chaotropic
 PT agent and purification. after absorption to hydrophobic resin.
 XX
 PS Example 1; Page 7; 16pp; German.
 XX
 CC The present sequence is an example of a proinsulin molecule corresp. to
 CC the general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In formula
 CC (II), X = Lys, Arg or a peptide of 2-35 amino acids contg. Lys or Arg at
 CC the N- and C-termini; Y = a natural amino acid; R1 = Phe or Arg; R2 =
 CC H, Arg, Lys, a peptide of 2-45 amino acids contg. Arg or Lys at the N-
 CC and C-termini; R3 = a natural amino acid; (A2-A20) and (B2-B29) are the
 CC insulin A- and B-chain sequences from human or other insulin. The
 CC proinsulin molecule (produced in recombinant E.coli) is reacted with
 CC mercaptan at a ratio of 2-10 SH residues of mercaptan per Cys residue of
 CC proinsulin. The reaction takes place in the presence of a chaotropic
 CC auxiliary agent at pH 10-11 and results in proinsulin with correctly
 CC linked cysteine bridges. Reaction with trypsin and opt. carboxypeptidase B
 CC yields correctly folded insulin. The insulin is isolated by absorption on
 CC a hydrophobic resin
 XX
 SO Sequence 97 AA;
 Query Match 98.1%; Score 504; DB 2; Length 97;
 Best Local Similarity 99.0%; Pred. No. 1.2e-47;
 Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 ATGTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADPOVQVELGGGPGA 60
 Db 2 ATGTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADPOVQVELGGGPGA 61
 Oy 61 GSLOPLALGSLQKRGIVEQCCTSIQSLYOLENYCG 96
 Db 62 GSLOPLALGSLQKRGIVEQCCTSIQSLYOLENYCG 97
 RESULT 6
 AAR15472
 ID AAR15472 standard; protein; 97 AA.
 XX
 AC AAR15472;
 XX
 DT 09-JAN-2003 (revised)
 DT 11-MAR-1992 (first entry)
 XX
 DE NcoI-sall fragment prod. as foreign protein.
 XX
 KW Signal peptide; promoter; transcription control; protease; IFN; IL; EGF;
 KM GM-CSF; Factor VIII; Insulin; TNF; NGF; vector; foreign protein.
 XX
 OS Brevibacillus brevis.
 XX
 OS Synthetic.
 XX
 PN WO9118101-A.
 XX
 PD 28-NOV-1991.
 XX
 PF 11-MAY-1990; 90JP-00122166.
 XX
 PR 11-MAY-1990; 90JP-00122166.
 PR 30-NOV-1990; 90JP-00334575.
 XX
 XX (FARH) HOECHST JAPAN LTD.
 XX

XX Hashimoto T, Tsujimura A, Udaka J;
 PI
 CC
 DR WPI; 1991-369253/50.
 DR N-PSDB; AA015230.
 PT Expression vector for signal peptide from *Bacillus brevis* - for
 PT expression of interferon, insulin, epidermal growth factor etc. with
 PT improved yield.
 XX
 PS Disclosure; Fig 9; 44pp; Japanese.
 CC
 CC DNA encoding the signal peptide associated with the BBRP42 gene of *B.*
 CC *brevis* may be incorporated into an expression vector, pref. with the
 CC promoter sequence and transcription control region. DNA encoding a
 CC foreign protein (e.g. EGF) may be ligated immediately downstream of the
 CC signal region. The vector may then be used to transform a host organism
 CC which then expresses the protein on culture. BBRP42 is secreted by *B.*
 CC *brevis* during an early stage of incubation, and so its associated signal
 CC peptide allows a foreign protein to be secreted during this early stage
 CC when protease prodn. is low, thus increasing recovery of the protein.
 CC Protein for which the method may be used include interferon,
 CC interleukin, human epidermal growth factor, GM-CSF, Factor VIII, insulin,
 CC tumour necrosis factor and NGF. See also AA015229-31. (Updated on 09-JAN-
 CC 2003 to add missing OS field.)
 XX
 SQ Sequence 97 AA;

Query Match 97.5%; Score 501; DB 2; Length 97;
 Best Local Similarity 98.9%; Pred. No. 2.5e-47;
 Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRRAEDDPQGVGVELGGGPGA 60
 |||||
 DB 2 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRRAEDDPQGVGVELGGGPGA 61
 |||||
 QY 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYC 95
 |||||
 DB 62 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYC 96
 |||||

RESULT 7
 ID AAY08004 standard; peptide; 96 AA.
 XX
 AC AAY08004;

DT 20-MAR-2003 (revised)
 DT 08-JUL-1999 (first entry)

DE Human proinsulin protein fragment 1.

KW Insulin; human; A chain; B chain; C chain; pro-insulin; chaotropic agent.

OS Homo sapiens.

PN EP906918-A2.

PD 07-APR-1999.

PF 11-AUG-1998; 98EP-00115048.

PR 18-AUG-1997; 97DE-01035711.

PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.

PI Rubroeder F, Keller R;

DR WPI; 1999-135401/22.

PT Preparation of insulin precursors with cystine bridges in the presence of
 PT cysteine and chaotropic auxiliary.
 XX

PS Example 1; Page 12; 16pp; German.

XX This invention describes a method for the production of insulin or
 CC insulin derivative precursors in an aqueous suspension which are diluted
 CC with cysteine or cysteine hydrochloride to pH 11.5 and then at 55 deg. C
 CC a chaotropic additive is added. This sequence represents human pro-
 CC insulin which is used in the description of the invention. This patent is
 CC an equivalent to HU9801886. (Updated on 20-MAR-2003 to correct DR field.)
 XX

SQ Sequence 96 AA;

Query Match 96.9%; Score 498; DB 2; Length 96;
 Best Local Similarity 98.9%; Pred. No. 5.2e-47;
 Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRRAEDDPQGVGVELGGGPGA 60
 |||||
 DB 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRRAEDDPQGVGVELGGGPGA 60
 |||||
 QY 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYC 95
 |||||
 DB 61 GSIQPLALEGSLQKRGIVGECCTSIQSLYQLENYC 95
 |||||

RESULT 8
 ID AAO17830 standard; peptide; 96 AA.
 XX
 AC AAO17830;

DT 13-SEP-2002 (first entry)

DE Human proinsulin #1.

KW Human; insulin; cysteine bridge; kaotropic additive.

OS Homo sapiens.

PN EP980874-A1.

PD 23-FEB-2000.

PF 11-AUG-1998; 99EP-00115386.

PR 18-AUG-1997; 97DE-01035711.

PR 11-AUG-1998; 98EP-00115048.

PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.

PI Rubroeder F, Keller R;

DR WPI; 1999-135401/22.

PT Preparation of insulin precursors with cystine bridges in the presence of
 PT cysteine and chaotropic auxiliary.

PS Example 1; Page 6; 20pp; German.

CC The present invention relates to insulin or insulin derivative precursors
 CC in an aqueous suspension, which are diluted with cysteine or cysteine
 CC hydrochloride to pH 11.5 and then at 55plusec a kaotropic additive is
 CC added. The present sequence is one version of human proinsulin
 XX

SQ Sequence 96 AA;

Query Match 96.9%; Score 498; DB 2; Length 96;
 Best Local Similarity 98.9%; Pred. No. 5.2e-47;
 Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRRAEDDPQGVGVELGGGPGA 60
 |||||
 DB 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRRAEDDPQGVGVELGGGPGA 60
 |||||

Qy 61 GSLOPLALEGSLQKRGIVEOCCTSSICSLYOLENYC 95
 |||||
 Db 61 GSLOPLALEGSLQKRGIVEOCCTSSICSLYOLENYC 95

RESULT 9

AA68898 standard; peptide; 97 AA.
 ID AAR68898
 XX
 AC AAR68898;
 XX
 DT 25-MAR-2003 (revised)
 DT 02-MAR-1995 (first entry)
 XX
 DE Human pro-insulin 1.
 XX
 KW pro-insulin; A-chain; B-chain; C-chain; disulphide; mercaptan;
 KW chototropic agent.
 XX
 OS Homo sapiens.
 XX
 PN EP600372-A1.
 XX
 PD 08-JUN-1994.
 XX
 PP 25-NOV-1993; 93EP-00118993.
 XX
 PR 02-DEC-1992; 92DE-04240420.
 XX
 PA (FARH) HOECHST AG.
 XX
 PI Obermeier R, Gerl M, Ludwig J, Sabel W;
 XX
 DR WPI; 1994-177718/22.
 XX
 PT prodn. of pro-insulin with correct di: sulphide bridges - by treating
 PT recombinant precursor protein with mercaptan in alkali and in presence of
 PT chototropic agent, then isolation on hydrophobic resin.
 XX
 PS Disclosure; Page 10-11; 15pp; German.
 XX
 CC pro-insulin is produced by treating recombinant precursor protein with a
 CC mercaptan to provide 2-10 SH residues per Cys residue, in presence of a
 CC chaotropic agent and in aq. medium of pH 10-11, treating the prod. with 3
 CC -50 g hydrophobic adsorber resin per 1 aq. medium of pH 4-7, isolating
 CC the adsorbed resin and pro-insulin and desorbing the pro-insulin. This
 CC method produces pro-insulin with correctly bonded Cys bridges. Compared
 CC with known methods it involves fewer stages (esp. no sulphiteolysis or
 CC cyanogen bromide cleavage) and overall losses during purification are
 CC reduced, i.e. the process is quicker and gives better yields. Sequences
 CC of insulin chain A, B and C are given in AAR68895-97. Sequences of pro-
 CC insulin 1-4 are given in AAR68898-901. (Updated on 25-MAR-2003 to correct
 CC PN field.)
 XX
 CC

Sequence 97 AA;
 SQ

Query Match 96.9%; Score 498; DB 2; Length 97;
 Best Local Similarity 98.9%; Pred. No. 5.3e-47;
 Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVGGGPGA 60
 |||||
 Db 2 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVGGGPGA 61

Qy 61 GSLOPLALEGSLQKRGIVEOCCTSSICSLYOLENYC 95
 |||||
 Db 62 GSLOPLALEGSLQKRGIVEOCCTSSICSLYOLENYC 96

RESULT 10
 AAW93414
 ID AAW93414 standard; protein; 97 AA.
 XX

AC AAW93414;
 XX
 DT 11-JUN-1999 (first entry)
 XX
 DE Human insulin derivative peptide #6.
 XX
 KW Insulin derivative; human; treatment; diabetes.
 XX
 OS Synthetic.
 XX
 PN EP885961-A1.
 XX
 PD 23-DEC-1998.
 XX
 PP 15-JUN-1998; 98EP-00110889.
 XX
 PR 20-JUN-1997; 97DE-01026167.
 XX
 PA (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.
 XX
 PI Ertl J, Habermann P, Geisen K, Seipke G;
 XX
 DR WPI; 1999-047558/05.
 XX
 PP New fast acting insulin variants - with aminoacid substitutions at
 PT positions B3, B27, B28 and B29.
 XX
 PS Claim 38; Page 17-18; 28pp; German.
 XX
 CC This invention describes novel fast acting insulin derivatives which have
 CC a Asn residue at position B3 and B27-B29 replaced with another naturally
 CC occurring neutral or acidic amino acid, and with Asn at position A21
 CC optionally replaced by Asp, Gly, Ser, Thr or Ala, and Phe at position B1,
 CC and the amino acid at position B1 are optionally deleted. Such
 CC derivatives can be used for treating diabetes. The derivatives have a
 CC faster onset of action than that of wild-type human insulin, especially
 CC when administered subcutaneously
 XX
 CC

Sequence 97 AA;
 SQ

Query Match 96.9%; Score 498; DB 2; Length 97;
 Best Local Similarity 97.9%; Pred. No. 5.3e-47;
 Matches 93; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVGGGPGA 60
 |||||
 Db 2 ATSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVGGGPGA 61

Qy 61 GSLOPLALEGSLQKRGIVEOCCTSSICSLYOLENYC 95
 |||||
 Db 62 GSLOPLALEGSLQKRGIVEOCCTSSICSLYOLENYC 96

RESULT 11

AA51222
 ID AA51222 standard; protein; 98 AA.
 XX
 AC AA51222;
 XX
 DT 06-APR-2000 (first entry)
 XX
 DE Human preproinsulin B-chain protein analog His(B0).
 XX
 KW insulin; human; antidiabetic; zinc; diabetes mellitus; treatment;
 KW pharmacodynamic; analog; B chain.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN DE19825447-A1.
 XX
 PP 09-DEC-1999.
 XX

```

XX 06-JUN-1998; 98DE-01025447.
XX
XX 06-JUN-1998; 98DE-01025447.
XX
PA (HMRI ) HOECHST MARION ROUSSEL DEUT GMBH.
XX
PI Ertl J, Habermann P, Geisen K, Seipke G, Wollmer A;
XX WPI; 2000-098618/09.
XX
PT New insulin analogs and their zinc complexes, with slow onset of
XX activity, for treating diabetes mellitus.
XX
PS Claim 27; Page 18; 22pp; German.
XX
CC This invention describes novel human insulin analogs (I) and their salts.
CC which have antidiabetic activity. (I), also their complexes with zinc,
CC are used for treatment of diabetes mellitus, types I or II. (I) have
CC increased zinc-binding capacity, form stable complexes with zinc and
CC compared with native human insulin have a delayed onset of activity
CC profile, so provide better control of diabetes following subcutaneous
CC injection. Zinc-free formulations do not have slow-release properties, so
CC variation of the zinc content provides control of pharmacodynamics (not
CC possible with native human insulin), e.g. tailored to the requirements of
CC individual patients. This sequence represents a human preproinsulin B
CC chain analog which is used in the method of the invention
XX
SQ Sequence 98 AA;
XX
Query Match 96.8%; Score 497.5; DB 3; Length 98;
Best Local Similarity 99.0%; Pred. No. 6.1e-47;
Matches 95; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
XX
QY 1 ATTSNGNSAR--FVNQHLGSHLYEALYVCGERGFFYPTKTRREADDPQGVQVELGGGP 59
   |||||
   2 ATTSNGNSARHFVNQHLGSHLYEALYVCGERGFFYPTKTRREADDPQGVQVELGGGP 61
XX
Db 60 AGSLQPLALEGSLQKRGIVEOCCTSIQSLYQLENYC 95
   |||||
   62 AGSLQPLALEGSLQKRGIVEOCCTSIQSLYQLENYC 97
XX
Db
XX
RESULT 12
AAVS1223
ID AAVS1223 standard; protein; 99 AA.
XX
AC AAVS1223;
XX
DT 06-APR-2000 (first entry)
XX
DE Human preproinsulin B-chain protein analog His(B-1).
XX
KW Insulin; human; antidiabetic; zinc; diabetes mellitus; treatment;
XX pharmacodynamic; analog; B chain.
XX
OS Synthetic.
XX OS Homo sapiens.
XX
PN DE19825447-A1.
XX
PD 09-DEC-1999.
XX
PF 06-JUN-1998; 98DE-01025447.
XX
PR 06-JUN-1998; 98DE-01025447.
XX
PA (HMRI ) HOECHST MARION ROUSSEL DEUT GMBH.
XX
PI Ertl J, Habermann P, Geisen K, Seipke G, Wollmer A;
XX WPI; 2000-098618/09.
XX

```

```

PT New insulin analogs and their zinc complexes, with slow onset of
XX activity, for treating diabetes mellitus.
XX
PS Claim 28; Page 19; 22pp; German.
XX
CC This invention describes novel human insulin analogs (I) and their salts.
CC which have antidiabetic activity. (I), also their complexes with zinc,
CC are used for treatment of diabetes mellitus, types I or II. (I) have
CC increased zinc-binding capacity, form stable complexes with zinc and
CC compared with native human insulin have a delayed onset of activity
CC profile, so provide better control of diabetes following subcutaneous
CC injection. Zinc-free formulations do not have slow-release properties, so
CC variation of the zinc content provides control of pharmacodynamics (not
CC possible with native human insulin), e.g. tailored to the requirements of
CC individual patients. This sequence represents a human preproinsulin B
CC chain analog which is used in the method of the invention
XX
SQ Sequence 99 AA;
XX
Query Match 96.7%; Score 497; DB 3; Length 99;
Best Local Similarity 97.9%; Pred. No. 7e-47;
Matches 95; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
XX
QY 1 ATTSNGNSAR--FVNQHLGSHLYEALYVCGERGFFYPTKTRREADDPQGVQVELGGP 58
   |||||
   2 ATTSNGNSARHAFVNQHLGSHLYEALYVCGERGFFYPTKTRREADDPQGVQVELGGP 61
XX
Db 59 GAGSLQPLALEGSLQKRGIVEOCCTSIQSLYQLENYC 95
   |||||
   62 GAGSLQPLALEGSLQKRGIVEOCCTSIQSLYQLENYC 98
XX
Db
XX
RESULT 13
AAVS1224
ID AAVS1224 standard; protein; 100 AA.
XX
AC AAVS1224;
XX
DT 06-APR-2000 (first entry)
XX
DE Human preproinsulin B-chain analog His(B-2), Ala(B-1), Ala(B0).
XX
KW Insulin; human; antidiabetic; zinc; diabetes mellitus; treatment;
XX pharmacodynamic; analog; B chain.
XX
OS Synthetic.
XX OS Homo sapiens.
XX
PN DE19825447-A1.
XX
PD 09-DEC-1999.
XX
PF 06-JUN-1998; 98DE-01025447.
XX
PR 06-JUN-1998; 98DE-01025447.
XX
PA (HMRI ) HOECHST MARION ROUSSEL DEUT GMBH.
XX
PI Ertl J, Habermann P, Geisen K, Seipke G, Wollmer A;
XX WPI; 2000-098618/09.
XX
PT New insulin analogs and their zinc complexes, with slow onset of
XX activity, for treating diabetes mellitus.
XX
PS Claim 29; Page 20; 22pp; German.
XX
CC This invention describes novel human insulin analogs (I) and their salts.
CC which have antidiabetic activity. (I), also their complexes with zinc,
CC are used for treatment of diabetes mellitus, types I or II. (I) have
CC increased zinc-binding capacity, form stable complexes with zinc and
CC compared with native human insulin have a delayed onset of activity
CC profile, so provide better control of diabetes following subcutaneous
CC

```

CC injection. Zinc-free formulations do not have slow-release properties, so
CC variation of the zinc content provides control of pharmacodynamics (not
CC possible with native human insulin), e.g. tailored to the requirements of
CC individual patients. This sequence represents a human preproinsulin B
CC chain analog which is used in the method of the invention
XX

50 Sequence 100 AA;

Query Match 96.6%; Score 496.5; DB 3; Length 100;

Best Local Similarity 97.9%; Pred. No. 8e-47; Indels 3; Gaps 1;
Matches 95; Conservative 0; Mismatches 0;

Oy 1 ATTSTGNSARFVNHLCGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLGGG 57
Db 2 ATTSTGNSARFVNHLCGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLGGG 61

Oy 58 PGAGSLQPLALEGSLQKRGIVEQCCTSIQSLYLENYC 95
Db 62 PGAGSLQPLALEGSLQKRGIVEQCCTSIQSLYLENYC 99

RESULT 14

AAW93416 AAW93416 standard; protein; 97 AA.

XX AC AAW93416;

DT 11-JUN-1999 (first entry)

DE Human insulin derivative peptide #8.

KW Insulin derivative; human; treatment; diabetes.

OS Synthetic.

OS Homo sapiens.

PN EP885961-A1.

PD 23-DEC-1998.

PP 15-JUN-1998; 98EP-00110889.

PR 20-JUN-1997; 97DE-01026167.

XX (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.

PI Ertl J, Habermann P, Geisen K, Seipke G;

XX WPI, 1999-047558/05.

PT New fast acting insulin variants - with aminoacid substitutions at
PT positions B3, B27, B28 and B29.

PS Claim 39; Page 19; 28pp; German.

CC This invention describes novel fast acting insulin derivatives which have
CC a Asn residue at position B3 and B27-B29 replaced with another naturally
CC occurring neutral or acidic amino acid, and with Asn at position A21
CC optionally replaced by Asp, Gly, Ser, Thr or Ala, and Phe at position B1,
CC and the amino acid at position B1 are optionally deleted. Such
CC derivatives can be used for treating diabetes. The derivatives have a
CC faster onset of action than that of wild-type human insulin, especially
CC when administered subcutaneously
XX

50 Sequence 97 AA;

Query Match 96.5%; Score 496; DB 2; Length 97;

Best Local Similarity 97.9%; Pred. No. 8.8e-47; Indels 0; Gaps 0;
Matches 93; Conservative 0; Mismatches 2;

Oy 1 ATTSTGNSARFVNHLCGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLGGG 60
Db 2 ATTSTGNSARFVNHLCGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLGGG 61

Oy 61 GSLOPLALEGSLQKRGIVEQCCTSIQSLYLENYC 95
Db 62 GSLOPLALEGSLQKRGIVEQCCTSIQSLYLENYC 96

RESULT 15

AAW93419 AAW93419 standard; protein; 97 AA.

XX AC AAW93419;

DT 11-JUN-1999 (first entry)

DE Human insulin derivative peptide #11.

KW Insulin derivative; human; treatment; diabetes.

OS Synthetic.

OS Homo sapiens.

PN EP885961-A1.

PD 23-DEC-1998.

PP 15-JUN-1998; 98EP-00110889.

PR 20-JUN-1997; 97DE-01026167.

XX (HMRI) HOECHST MARION ROUSSEL DEUT GMBH.

PI Ertl J, Habermann P, Geisen K, Seipke G;

XX WPI, 1999-047558/05.

PT New fast acting insulin variants - with aminoacid substitutions at
PT positions B3, B27, B28 and B29.

PS Claim 37; Page 20-21; 28pp; German.

CC This invention describes novel fast acting insulin derivatives which have
CC a Asn residue at position B3 and B27-B29 replaced with another naturally
CC occurring neutral or acidic amino acid, and with Asn at position A21
CC optionally replaced by Asp, Gly, Ser, Thr or Ala, and Phe at position B1,
CC and the amino acid at position B1 are optionally deleted. Such
CC derivatives can be used for treating diabetes. The derivatives have a
CC faster onset of action than that of wild-type human insulin, especially
CC when administered subcutaneously
XX

50 Sequence 97 AA;

Query Match 95.7%; Score 492; DB 2; Length 97;

Best Local Similarity 97.9%; Pred. No. 2.4e-46; Indels 0; Gaps 0;
Matches 93; Conservative 0; Mismatches 2;

Oy 1 ATTSTGNSARFVNHLCGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLGGG 60
Db 2 ATTSTGNSARFVNHLCGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLGGG 61

Oy 61 GSLOPLALEGSLQKRGIVEQCCTSIQSLYLENYC 95
Db 62 GSLOPLALEGSLQKRGIVEQCCTSIQSLYLENYC 96

Search completed: May 20, 2005, 00:41:40
Job time : 166 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 00:36:01 ; Search time 43 Seconds
(without alignments)
166.658 Million cell updates/sec

Title: US-10-632-414-3
Perfect score: 514
Sequence: 1 ATSTGNSARFVNQHLGSH.....IVEQCTSCISLYQLENYCG 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA.*
2: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/PTCUS_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	504	98.1	96	2	US-09-134-836-5
2	504	98.1	96	3	US-09-386-303A-5
3	504	98.1	96	4	US-09-947-563-5
4	504	98.1	97	1	US-08-389-487-7
5	498	96.9	96	2	US-09-134-836-4
6	498	96.9	96	3	US-09-386-303A-4
7	498	96.9	96	4	US-09-947-563-4
8	498	96.9	97	1	US-08-160-376A-4
9	498	96.9	97	3	US-09-099-307-6
10	497.5	96.8	98	4	US-09-701-968-7
11	497	96.7	99	4	US-09-701-968-8
12	496.5	96.6	100	4	US-09-701-968-9
13	496	96.5	97	3	US-09-099-307-8
14	492	95.7	97	3	US-09-099-307-7
15	492	95.7	97	3	US-09-099-307-11
16	479	93.2	90	1	US-08-030-731A-43
17	479	88.3	110	3	US-08-950-720A-11
18	454	88.3	110	3	US-08-589-028-2
19	454	88.3	110	3	US-08-784-582-2
20	454	88.3	110	3	US-08-785-271-2
21	454	88.3	110	4	US-08-472-701-2
22	454	88.3	110	4	US-09-185-852-2
23	454	88.3	110	4	US-09-815-239-3
24	454	88.3	110	4	US-09-617-389B-20
25	454	88.3	110	4	US-09-323-738-2
26	454	88.3	110	4	US-09-015-399-7
27	454	88.3	110	5	PCT-US95-08596-2

28	454	88.3	151	2	US-08-508-664-15	Sequence 15, Appl
29	454	88.3	161	2	US-08-508-664-16	Sequence 16, Appl
30	453	88.1	117	4	US-09-280-030-63	Sequence 63, Appl
31	453	88.1	130	4	US-09-280-030-62	Sequence 62, Appl
32	450	87.5	86	4	US-09-477-924-2	Sequence 2, Appl
33	450	87.5	86	4	US-09-723-981-2	Sequence 2, Appl
34	450	87.5	86	4	US-09-723-981-2	Sequence 2, Appl
35	450	87.5	86	4	US-09-723-981-2	Sequence 2, Appl
36	450	87.5	167	1	US-07-918-953-8	Sequence 8, Appl
37	450	87.5	167	1	US-08-081-661-8	Sequence 8, Appl
38	440	85.6	91	4	US-09-676-787-7	Sequence 7, Appl
39	440	85.6	110	4	US-09-574-443-1	Sequence 1, Appl
40	435	84.6	110	3	US-08-589-028-4	Sequence 4, Appl
41	435	84.6	110	3	US-08-784-582-4	Sequence 4, Appl
42	435	84.6	110	3	US-08-785-271-4	Sequence 4, Appl
43	316	61.5	65	3	US-08-900-574-3	Sequence 3, Appl
44	316	61.5	67	3	US-08-900-574-7	Sequence 7, Appl
45	315.5	61.4	66	3	US-08-900-574-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-09-134-836-5
: Sequence 5, Application US/09134836
: Patent No. 5986048
:
: GENERAL INFORMATION:
: APPLICANT: Rubroder, Franz-Josef
: APPLICANT: Keller, Reinhold
: TITLE OF INVENTION: Improved process for obtaining
: TITLE OF INVENTION: Insulin precursors having correctly bonded cysteine bridges
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Farnegun, Henderson, Farrabow, Garrett &
: ADDRESSEE: Dunner
: STREET: 1300 I Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-3315
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/134,836
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: Leslie McDonnell
: REGISTRATION NUMBER: 34,872
: REFERENCE/DOCKET NUMBER: 02481.1600-00000
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 408-4000
: TELEFAX: (202) 408-4400
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 96 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORGANISM: Escherichia coli
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..96
: US-09-134-836-5
:
: Query Match 98.1%; Score 504; DB 2; Length 96;
: Best Local Similarity 99.0%; Pred. No. 7,7e-52;
```



```

APPLICANT: Obermeier, Rainer
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jurgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process for Obtaining Insulin Having
TITLE OF INVENTION: Correctly Linked Cysteine Bridges
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/389,487
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Einaudi, Carol P.
REGISTRATION NUMBER: 32,220
REFERENCE/DOCKET NUMBER: 02481.1424-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4000
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-389-487-7

Query Match          98.1%; Score 504; DB 1; Length 97;
Best Local Similarity 99.0%; Pred. No. 7.8e-52;
Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTGNSARFVNQHLGSHLYEALYLVCGRGFFYTKTRREADPOVQVLEGGGPGA 60
DB 2 ATTGNSARFVNQHLGSHLYEALYLVCGRGFFYTKTRREADPOVQVLEGGGPGA 61

QY 61 GSIQPLALEGSLQKRGIVEOCCTSIQSLYQLENYCG 96
DB 62 GSIQPLALEGSLQKRGIVEOCCTSIQSLYQLENYCG 97

RESULT 5
US-09-134-836-4
Sequence 4, Application US/09134836
Patent No. 5986048
GENERAL INFORMATION:
APPLICANT: Rubroder, Franz-Josef
APPLICANT: Keller, Reinhold
TITLE OF INVENTION: Improved process for obtaining
TITLE OF INVENTION: Insulin precursors having correctly bonded cysteine bridges
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/134,836
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDonnell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
US-09-134-836-4

Query Match          96.9%; Score 498; DB 2; Length 96;
Best Local Similarity 98.9%; Pred. No. 3.9e-51;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTGNSARFVNQHLGSHLYEALYLVCGRGFFYTKTRREADPOVQVLEGGGPGA 60
DB 1 ATTGNSARFVNQHLGSHLYEALYLVCGRGFFYTKTRREADPOVQVLEGGGPGA 60

QY 61 GSIQPLALEGSLQKRGIVEOCCTSIQSLYQLENYC 95
DB 61 GSIQPLALEGSLQKRGIVEOCCTSIQSLYQLENYC 95

RESULT 6
US-09-386-303A-4
Sequence 4, Application US/09386303A
Patent No. 6380355
GENERAL INFORMATION:
APPLICANT: Rubroder, Franz-Josef
APPLICANT: Keller, Reinhold
TITLE OF INVENTION: Improved process for obtaining
TITLE OF INVENTION: Insulin precursors having correctly bonded cysteine bridges
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESS: Finnegan, Henderson, Farabow, Garrett &
ADDRESS: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/386,303A
FILING DATE: 31-Aug-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDonnell

```

REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-386-303A-4

Query Match 96.9%; Score 498; DB 3; Length 96;
Best Local Similarity 98.9%; Pred. No. 3.9e-51;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRRAEDDPVGVVELGGGPGA 60
DB 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRRAEDLPVGVVELGGGPGA 60

QY 61 GSIQPLALEGSLQKRGIVQCCTSIQSLVQLENYC 95
DB 61 GSIQPLALEGSLQKRGIVQCCTSIQSLVQLENYC 95

RESULT 7

US-947-563-4
Sequence 4, Application US/09947563
Patent No. 6727346
GENERAL INFORMATION:
APPLICANT: Rudröder, Franz-Josef
Keller, Reinhold
TITLE OF INVENTION: Improved process for obtaining
insulin precursors having correctly bonded cysteine bridges
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fünnegan, Henderson, Farrabow, Garrett &
Dunnet
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3115
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/947,563
FILING DATE: 07-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDowell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4

Query Match 96.9%; Score 498; DB 4; Length 96;
Best Local Similarity 98.9%; Pred. No. 3.9e-51;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRRAEDDPVGVVELGGGPGA 60
DB 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRRAEDLPVGVVELGGGPGA 60

QY 61 GSIQPLALEGSLQKRGIVQCCTSIQSLVQLENYC 95
DB 61 GSIQPLALEGSLQKRGIVQCCTSIQSLVQLENYC 95

RESULT 8

US-08-160-376A-4
Sequence 4, Application US/08160376A
Patent No. 5473049
GENERAL INFORMATION:
APPLICANT: Obermeier, Ranier
APPLICANT: Gerl, Martin
APPLICANT: Ludwig, Jürgen
APPLICANT: Sabel, Walter
TITLE OF INVENTION: Process For Obtaining Proinsulin
TITLE OF INVENTION: Possessing correctly linked
cysteine bridges
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenneth A. Genoni, Esq.
STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
CITY: Somerville
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08876-1258
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE
COMPUTER: IBM 386
OPERATING SYSTEM: WINDOWS 3.1
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/160,376A
FILING DATE: December 1, 1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GE P 4240420.7
FILING DATE: December 2, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Barbara V. Maurer, Esq.
REGISTRATION NUMBER: 31,287
REFERENCE/DOCKET NUMBER: HOE 92/F 384
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 231-4079
TELEFAX: (908) 231-2255
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 Amino Acids
TYPE: Amino Acid (AA)
TOPOLOGY: not relevant
US-08-160-376A-4

Query Match 96.9%; Score 498; DB 1; Length 97;
Best Local Similarity 98.9%; Pred. No. 3.9e-51;

Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGGA 60
 |||||
 Db 2 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGGA 61
 |||||

Qy 61 GSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 95
 |||||
 Db 62 GSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 96
 |||||

RESULT 9
 US-09-099-307-6
 ; Sequence 6, Application US/09099307A
 ; Patent No. 6221633
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; TITLE OF INVENTION: NOVEL INSULIN DERIVATIVES HAVING A RAPID ONSET OF ACTION
 ; FILE REFERENCE: 02481.1597-00000
 ; CURRENT APPLICATION NUMBER: US/09/099.307A
 ; CURRENT FILING DATE: 1998-06-18
 ; EARLIER APPLICATION NUMBER: 19726167.1
 ; EARLIER FILING DATE: 1997-06-20
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 6
 ; LENGTH: 97
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-099-307-6

Query Match 96.9%; Score 498; DB 3; Length 97;
 Best Local Similarity 97.9%; Pred. No. 3.9e-51;
 Matches 93; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGGA 60
 |||||
 Db 2 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGGA 61
 |||||

Qy 61 GSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 95
 |||||
 Db 62 GSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 96
 |||||

RESULT 10
 US-09-701-968-7
 ; Sequence 7, Application US/09701968
 ; Patent No. 6686177
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; APPLICANT: WOLLMER, AXEL
 ; TITLE OF INVENTION: NOVEL INSULIN ANALOGS WITH ENHANCED ZINC BINDING
 ; FILE REFERENCE: 02481.1719-00000
 ; CURRENT APPLICATION NUMBER: US/09/701.968
 ; CURRENT FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 98
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-701-968-7

Query Match 96.8%; Score 497.5; DB 4; Length 98;
 Best Local Similarity 99.0%; Pred. No. 4.6e-51;
 Matches 95; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGPG 59
 |||||
 Db 2 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGPG 61
 |||||

Qy 60 AGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 95
 |||||
 Db 62 AGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 97
 |||||

RESULT 11
 US-09-701-968-8
 ; Sequence 8, Application US/09701968
 ; Patent No. 6686177
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; APPLICANT: WOLLMER, AXEL
 ; TITLE OF INVENTION: NOVEL INSULIN ANALOGS WITH ENHANCED ZINC BINDING
 ; FILE REFERENCE: 02481.1719-00000
 ; CURRENT APPLICATION NUMBER: US/09/701.968
 ; CURRENT FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 8
 ; LENGTH: 99
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-701-968-8

Query Match 96.7%; Score 497; DB 4; Length 99;
 Best Local Similarity 97.9%; Pred. No. 5.3e-51;
 Matches 95; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGP 58
 |||||
 Db 2 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGGP 61
 |||||

Qy 59 GAGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 95
 |||||
 Db 62 GAGSLQPLALEGSLQKRGIVEQCCTSIQSLYQLENYC 98
 |||||

RESULT 12
 US-09-701-968-9
 ; Sequence 9, Application US/09701968
 ; Patent No. 6686177
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; APPLICANT: WOLLMER, AXEL
 ; TITLE OF INVENTION: NOVEL INSULIN ANALOGS WITH ENHANCED ZINC BINDING
 ; FILE REFERENCE: 02481.1719-00000
 ; CURRENT APPLICATION NUMBER: US/09/701.968
 ; CURRENT FILING DATE: 2000-12-06
 ; NUMBER OF SEQ ID NOS: 15
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 100
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-701-968-9

Query Match 96.6%; Score 496.5; DB 4; Length 100;
 Best Local Similarity 96.9%; Pred. No. 6.2e-51;
 Matches 95; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGG 57
 |||||
 Db 2 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREADDPQGVGVELGGG 61
 |||||

OY 58 PGAGSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 95
 |||||
 DB 62 PGAGSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 99

RESULT 13

US-09-099-307-8
 ; Sequence 8, Application US/09099307A
 ; Patent No. 6221633
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; TITLE OF INVENTION: NOVEL INSULIN DERIVATIVES HAVING A RAPID ONSET OF ACTION
 ; FILE REFERENCE: 02481.1597-00000
 ; CURRENT APPLICATION NUMBER: US/09/099,307A
 ; CURRENT FILING DATE: 1998-06-18
 ; EARLIER APPLICATION NUMBER: 19726167.1
 ; EARLIER FILING DATE: 1997-06-20
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 8
 ; LENGTH: 97
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-099-307-8

Query Match 96.5%; Score 496; DB 3; Length 97;
 Best Local Similarity 97.9%; Pred. No. 6,8e-51;
 Matches 93; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ATTSTGNSARFVNOHLCGSHLVEALYLVGGERGFYTPKTRREADDPQGVGVELGGGPGA 60
 |||||
 DB 2 ATTSTGNSARFVNOHLCGSHLVEALYLVGGERGFYTPKTRREADDPQGVGVELGGGPGA 61
 |||||
 OY 61 GSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 95
 |||||
 DB 62 GSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 96
 |||||

RESULT 14

US-09-099-307-7
 ; Sequence 7, Application US/09099307A
 ; Patent No. 6221633
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; TITLE OF INVENTION: NOVEL INSULIN DERIVATIVES HAVING A RAPID ONSET OF ACTION
 ; FILE REFERENCE: 02481.1597-00000
 ; CURRENT APPLICATION NUMBER: US/09/099,307A
 ; CURRENT FILING DATE: 1998-06-18
 ; EARLIER APPLICATION NUMBER: 19726167.1
 ; EARLIER FILING DATE: 1997-06-20
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 97
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-099-307-7

Query Match 95.7%; Score 492; DB 3; Length 97;
 Best Local Similarity 97.9%; Pred. No. 2e-50;
 Matches 93; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ATTSTGNSARFVNOHLCGSHLVEALYLVGGERGFYTPKTRREADDPQGVGVELGGGPGA 60
 |||||
 DB 2 ATTSTGNSARFVNOHLCGSHLVEALYLVGGERGFYTPKTRREADDPQGVGVELGGGPGA 61
 |||||

OY 61 GSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 95
 |||||
 DB 62 GSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 96
 |||||

RESULT 15

US-09-099-307-11
 ; Sequence 11, Application US/09099307A
 ; Patent No. 6221633
 ; GENERAL INFORMATION:
 ; APPLICANT: ERTL, JOHANN
 ; APPLICANT: HABERMANN, PAUL
 ; APPLICANT: GEISEN, KARL
 ; APPLICANT: SEIPKE, GERHARD
 ; TITLE OF INVENTION: NOVEL INSULIN DERIVATIVES HAVING A RAPID ONSET OF ACTION
 ; FILE REFERENCE: 02481.1597-00000
 ; CURRENT APPLICATION NUMBER: US/09/099,307A
 ; CURRENT FILING DATE: 1998-06-18
 ; EARLIER APPLICATION NUMBER: 19726167.1
 ; EARLIER FILING DATE: 1997-06-20
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 11
 ; LENGTH: 97
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-099-307-11

Query Match 95.7%; Score 492; DB 3; Length 97;
 Best Local Similarity 97.9%; Pred. No. 2e-50;
 Matches 93; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ATTSTGNSARFVNOHLCGSHLVEALYLVGGERGFYTPKTRREADDPQGVGVELGGGPGA 60
 |||||
 DB 2 ATTSTGNSARFVNOHLCGSHLVEALYLVGGERGFYTPKTRREADDPQGVGVELGGGPGA 61
 |||||
 OY 61 GSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 95
 |||||
 DB 62 GSLQPLALEGSLQKRGIVEQCCTSI CSLYQLENYC 96
 |||||

Search completed: May 20, 2005, 00:45:32
 Job time : 45 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 00:33:35 ; Search time 135 Seconds
(without alignments)
237.872 Million cell updates/sec

Title: US-10-632-414-3

Perfect score: 514

Sequence: 1 ATTGTGSAFVNQHLGSH.....IVEQCCTSGSLVLENYCG 96

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1434725 seqs, 334507595 residues

Total number of hits satisfying chosen parameters: 1434725

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications_MA:*

2: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*

4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*

6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*

8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*

10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*

11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*

12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*

13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*

14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*

15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*

16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*

17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*

18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*

19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*

20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	514	100.0	96 17 US-10-632-414-3	Sequence 3, Appl1
2	508	98.8	96 17 US-10-632-414-2	Sequence 2, Appl1
3	504	98.1	96 9 US-09-947-563-5	Sequence 5, Appl1
4	498	96.9	96 9 US-09-947-563-4	Sequence 4, Appl1
5	460	89.5	110 17 US-10-869-040-21	Sequence 21, Appl1
6	460	89.5	110 17 US-10-869-040-22	Sequence 22, Appl1
7	456	88.7	257 17 US-10-869-040-196	Sequence 196, App
8	454	88.3	110 9 US-09-205-658-125	Sequence 125, App
9	454	88.3	110 9 US-09-815-229-3	Sequence 3, Appl1
10	454	88.3	110 9 US-09-804-409A-9	Sequence 9, Appl1
11	454	88.3	110 10 US-09-969-748C-6	Sequence 6, Appl1
12	454	88.3	110 10 US-09-963-693-125	Sequence 125, App
13	454	88.3	110 14 US-10-038-686-1	Sequence 1, Appl1

14	454	88.3	110 14 US-10-328-813-2	Sequence 2, Appl1
15	454	88.3	110 15 US-10-346-563-2	Sequence 2, Appl1
16	454	88.3	110 15 US-10-321-117-2	Sequence 2, Appl1
17	454	88.3	110 15 US-10-411-037-44	Sequence 44, Appl1
18	454	88.3	110 15 US-10-411-026-44	Sequence 44, Appl1
19	454	88.3	110 15 US-10-410-962-44	Sequence 44, Appl1
20	454	88.3	110 15 US-10-411-049-44	Sequence 44, Appl1
21	454	88.3	110 15 US-10-700-725-20	Sequence 20, Appl1
22	454	88.3	110 16 US-10-410-930-44	Sequence 44, Appl1
23	454	88.3	110 16 US-10-410-997-44	Sequence 44, Appl1
24	454	88.3	110 16 US-10-411-012-44	Sequence 44, Appl1
25	454	88.3	110 16 US-10-287-994-44	Sequence 44, Appl1
26	454	88.3	110 16 US-10-740-098-3	Sequence 3, Appl1
27	454	88.3	110 16 US-10-410-913-44	Sequence 44, Appl1
28	454	88.3	110 17 US-10-410-980-44	Sequence 44, Appl1
29	454	88.3	110 17 US-10-869-040-7	Sequence 7, Appl1
30	454	88.3	110 17 US-10-869-040-26	Sequence 26, Appl1
31	454	88.3	110 17 US-10-410-897-44	Sequence 44, Appl1
32	453	88.1	117 9 US-10-280-030-63	Sequence 63, Appl1
33	453	88.1	130 9 US-09-280-030-62	Sequence 62, Appl1
34	451	87.7	110 15 US-10-383-285-2	Sequence 2, Appl1
35	451	87.7	110 17 US-10-869-040-23	Sequence 23, Appl1
36	450	87.5	86 9 US-09-878-380-1	Sequence 1, Appl1
37	450	87.5	86 10 US-09-858-935B-4	Sequence 2, Appl1
38	450	87.5	86 13 US-10-028-410-2	Sequence 2, Appl1
39	450	87.5	86 13 US-10-054-873-4	Sequence 2, Appl1
40	450	87.5	86 14 US-10-444-826-2	Sequence 4, Appl1
41	450	87.5	86 15 US-10-271-869-4	Sequence 4, Appl1
42	450	87.5	86 15 US-10-444-662-2	Sequence 2, Appl1
43	450	87.5	86 15 US-10-444-649-2	Sequence 2, Appl1
44	450	87.5	86 15 US-10-444-701-2	Sequence 2, Appl1
45	450	87.5	86 17 US-10-760-928-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-10-632-414-3

Sequence 3, Application US/10632414

Publication No. US2005008000A1

GENERAL INFORMATION:

APPLICANT: Aventis Pharma Deutschland GmbH

APPLICANT: THURLOW, Horst

APPLICANT: BLUMENSTOCK, Hans

APPLICANT: HAVENITH, Chantalie

APPLICANT: ERTL, Johann

TITLE OF INVENTION: A METHOD FOR PURIFYING PREPROINSULIN

FILE REFERENCE: DE4V2002/0053 US NP

CURRENT APPLICATION NUMBER: US/10/632,414

CURRENT FILING DATE: 2003-08-01

PRIOR APPLICATION NUMBER: 60/433,726

PRIOR FILING DATE: 2002-12-16

PRIOR APPLICATION NUMBER: DE10235168.6

PRIOR FILING DATE: 2002-08-01

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 3

LENGTH: 96

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Preproinsulin

OTHER INFORMATION: II

US-10-632-414-3

Query Match

Best Local Similarity 100.0%; Score 514; DB 17; Length 96;

Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATTGTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

11 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

21 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

31 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

41 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

51 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

61 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

71 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

81 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

91 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

101 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

111 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

121 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

131 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

141 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

151 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

161 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

171 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

181 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

191 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

201 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

211 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

221 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

231 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

241 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

251 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

261 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

271 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

281 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

291 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

301 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

311 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

321 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

331 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

341 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

351 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

361 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

371 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

381 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

391 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

401 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

411 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

421 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

431 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

441 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

451 ATTTGSAFVNQHLGSHLVEALYLVCGERGFFYPTKTRREAEDPQVGVELGGGPGA 60

Qy 61 GSIQPLALEGSLQKRGIVQCCTSIQSLYLENYCG 96
Db 61 GSIQPLALEGSLQKRGIVQCCTSIQSLYLENYCG 96

RESULT 2
US-10-632-414-2
; Sequence 2, Application US/10632414
; Publication No. US2005008000A1
; GENERAL INFORMATION:
; APPLICANT: Aventis Pharma Deutschland GmbH
; APPLICANT: THUROW, Horst
; APPLICANT: BLUMENSTOCK, Hans
; APPLICANT: HAVENITH, Chantalie
; APPLICANT: ERTL, Johann
; TITLE OF INVENTION: A METHOD FOR PURIFYING PREPROINSULIN
; FILE REFERENCE: DE42002/0053 US NP
; CURRENT APPLICATION NUMBER: US/10/632,414
; PRIOR FILING DATE: 2003-08-01
; PRIOR APPLICATION NUMBER: 60/433,726
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: DE10235168.6
; PRIOR FILING DATE: 2002-08-01
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Preproinsulin I
US-10-632-414-2

Query Match 98.8%; Score 508; DB 17; Length 96;
Best Local Similarity 100.0%; Pred. No. 2,1e-48;
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREADDPQGVGVLEGGGPGA 60
Db 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREADDPQGVGVLEGGGPGA 60
Qy 61 GSIQPLALEGSLQKRGIVQCCTSIQSLYLENYCG 95
Db 61 GSIQPLALEGSLQKRGIVQCCTSIQSLYLENYCG 95

RESULT 3
US-09-947-563-5
; Sequence 5, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulin precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/947,563
; FILING DATE: 07-Sep-2001

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/134,836
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Leslie McDowell
REGISTRATION NUMBER: 34,872
REFERENCE/DOCKET NUMBER: 02481.1600-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-947-563-5

Query Match 98.1%; Score 504; DB 9; Length 96;
Best Local Similarity 99.0%; Pred. No. 5,7e-48;
Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREADDPQGVGVLEGGGPGA 60
Db 1 ATTSTGNSARFVNQHLGSHLVEALYLVCGERGFFYPTKTRREADDPQGVGVLEGGGPGA 60
Qy 61 GSIQPLALEGSLQKRGIVQCCTSIQSLYLENYCG 96
Db 61 GSIQPLALEGSLQKRGIVQCCTSIQSLYLENYCG 96

RESULT 4
US-09-947-563-4
; Sequence 4, Application US/09947563
; Patent No. US20020156234A1
; GENERAL INFORMATION:
; APPLICANT: Rubroder, Franz-Josef
; APPLICANT: Keller, Reinhold
; TITLE OF INVENTION: Improved process for obtaining
; insulin precursors having correctly bonded cystine bridges
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farrahaw, Garrett &
; Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/947,563
; FILING DATE: 07-Sep-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/134,836
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Leslie McDowell
; REGISTRATION NUMBER: 34,872
; REFERENCE/DOCKET NUMBER: 02481.1600-00000


```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: Protein
LOCATION: 1..96
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-947-563-4

Query Match
Best Local Similarity 96.9%; Score 498; DB 9; Length 96;
Matches 94; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 1 ATGTGNSARFVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGA 60
Db 1 ATGTGNSARFVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGA 60
Cy 61 GSLOPLAEGSLQKRGIVECCTSIQSLYQLENYC 95
Db 61 GSLOPLAEGSLQKRGIVECCTSIQSLYQLENYC 95

RESULT 5
US-10-869-040-21
Sequence 21, Application US/10869040
Publication No. US20050039235A1
GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
APPLICANT: Boothe, Joseph
APPLICANT: Keon, Richard
APPLICANT: Nykiforuk, Cory
APPLICANT: Van Rooijen, Gijb
TITLE OF INVENTION: Methods for the Production of Insulin in Plants
FILE REFERENCE: 9369-301
CURRENT APPLICATION NUMBER: US/10/869,040
PRIOR FILING DATE: 2004-06-17
PRIOR APPLICATION NUMBER: 60/478,818
PRIOR FILING DATE: 2003-06-17
PRIOR APPLICATION NUMBER: 60/549,539
PRIOR FILING DATE: 2004-03-04
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 110
TYPE: PRT
ORGANISM: Macaca fascicularis
US-10-869-040-21

Query Match
Best Local Similarity 89.5%; Score 460; DB 17; Length 110;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 11 FVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGASLOPLAEG 70
Db 25 FVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGASLOPLAEG 84
Cy 71 SLOKRGIVECCTSIQSLYQLENYC 95
Db 85 SLOKRGIVECCTSIQSLYQLENYC 109

RESULT 6
US-10-869-040-22
Sequence 22, Application US/10869040
```

```
Publication No. US20050039235A1
GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
APPLICANT: Boothe, Joseph
APPLICANT: Keon, Richard
APPLICANT: Nykiforuk, Cory
APPLICANT: Van Rooijen, Gijb
TITLE OF INVENTION: Methods for the Production of Insulin in Plants
FILE REFERENCE: 9369-301
CURRENT APPLICATION NUMBER: US/10/869,040
PRIOR FILING DATE: 2004-06-17
PRIOR APPLICATION NUMBER: 60/478,818
PRIOR FILING DATE: 2003-06-17
PRIOR APPLICATION NUMBER: 60/549,539
PRIOR FILING DATE: 2004-03-04
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn version 3.1
SEQ ID NO 22
LENGTH: 110
TYPE: PRT
ORGANISM: Cercopithecus aethiops
US-10-869-040-22

Query Match
Best Local Similarity 89.5%; Score 460; DB 17; Length 110;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 11 FVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGASLOPLAEG 70
Db 25 FVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGASLOPLAEG 84
Cy 71 SLOKRGIVECCTSIQSLYQLENYC 95
Db 85 SLOKRGIVECCTSIQSLYQLENYC 109

RESULT 7
US-10-869-040-196
Sequence 196, Application US/10869040
Publication No. US20050039235A1
GENERAL INFORMATION:
APPLICANT: Moloney, Maurice M.
APPLICANT: Boothe, Joseph
APPLICANT: Keon, Richard
APPLICANT: Nykiforuk, Cory
APPLICANT: Van Rooijen, Gijb
TITLE OF INVENTION: Methods for the Production of Insulin in Plants
FILE REFERENCE: 9369-301
CURRENT APPLICATION NUMBER: US/10/869,040
PRIOR FILING DATE: 2004-06-17
PRIOR APPLICATION NUMBER: 60/478,818
PRIOR FILING DATE: 2003-06-17
PRIOR APPLICATION NUMBER: 60/549,539
PRIOR FILING DATE: 2004-03-04
NUMBER OF SEQ ID NOS: 196
SOFTWARE: PatentIn version 3.1
SEQ ID NO 196
LENGTH: 257
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Insulin fusion protein
US-10-869-040-196

Query Match
Best Local Similarity 88.7%; Score 456; DB 17; Length 257;
Matches 86; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Cy 3 TSTGNSARFVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGAS 62
Db 164 TRGGQHTTFVNQHLGSHLYVLCGERGFYTPKTRREADDPVGVVGLGGPGGAS 223
Cy 63 LQPLAEGSLQKRGIVECCTSIQSLYQLENYC 95
```

Query Match	88.3%;	Score 454;	DB 10;	Length 110;
Best Local Similarity	96.6%;	Pred. No. 2.3e-42;		
Matches 85; Conservative	1;	Mismatches 2;	Indels 0;	Gaps 0;

Qy	8	SARFVNQHLCSHLYEALYLVCGERGFYPTKTRREAEADPQVGQVEIGGGPAGSLQPLA	67
Db	22	AAAFVNQHLCSHLYEALYLVCGERGFYPTKTRREAEADPQVGQVEIGGGPAGSLQPLA	81
Qy	68	LEGLQKRGIVQECCTSLCYQLENTC	95
Db	82	LEGLQKRGIVQECCTSLCYQLENTC	109

```

RESULT 12
US-09-963-693-125
Sequence 125, Application US/09963693
Publication No. US20030181364A1
GENERAL INFORMATION:
APPLICANT: Ruvkun, Gary
APPLICANT: Ogg, Scott
TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
FILE REFERENCE: 00786/351004
CURRENT APPLICATION NUMBER: US/09/963,693
CURRENT FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US/09/205,658
PRIOR FILING DATE: 1998-12-03
PRIOR APPLICATION NUMBER: 08/857,076
PRIOR FILING DATE: 1997-05-15
PRIOR APPLICATION NUMBER: 08/888,534
PRIOR FILING DATE: 1997-07-07
PRIOR APPLICATION NUMBER: US96/10080
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 328
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 125
LENGTH: 110

```

ORGANISM: Homo sapiens
US-09-963-693-125

Query Match	88.3%	Score 454	DB 10	Length 110
Best Local Similarity	96.6%	Pred. NO. 2.3e-42		
Matches 85, Conservative	1	Mismatches 2	Indels 0	Gaps 0

OY 8 SARFVNQHLCSGSHLVEALYLVCGERGFFYTPTKTRREAEDDQVGGVLLGGSAGASLQPLA 67
 : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 22 AAAFVNQHLCGSHLVEALYLVCGERGFFYTPTKTRREAEDDQVGGVLLGGSAGASLQPLA 81

QY	68	LEGS	LQ	R	G	I	V	E	Q	C	T	S	I	C	S	L	Y	L	E	N	Y	C	95
Db	82	LEGS	LQ	R	G	I	V	E	Q	C	T	S	I	C	S	L	Y	L	E	N	Y	C	109

```

RESULT 13
US-10-038-686-1
: Sequence 1, Application US/10038686
: Publication No. US20030045467A1
: GENERAL INFORMATION:
: APPLICANT: Orban, Thamer
: TITLE OF INVENTION: AUTOANTIGEN VACCINE
: FILE REFERENCE: 10276-067001
: CURRENT APPLICATION NUMBER: US/10/038, 686
: CURRENT FILING DATE: 2002-05-31
: PRIOR APPLICATION NUMBER: 60/260, 068
: PRIOR FILING DATE: 2001-01-05
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 110
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-038-686-1

```

Query Match	88.3%	Score 454;	DB 14;	Length 110;
Best Local Similarity	96.6%	Pred. No. 2.3e-42;		
Matches	85;	Conservative	1;	Mismatches 2;
			Indels	0;
			Gaps	0;

Oy 8 SARPVNHLCGSHLVEALYLVCGERPFPTPKRRRAEDPQGVQLGGSPGAGSLQPLA 67
:
Db 22 AAAPVNHLTCSHLYEALYLVCGERPFPTPKRRRAEDLQGVQLGGSPGAGSLQPLA 81

Oy 68 LBSGLQRGTIVBOCCTSGLYOLENYC 95
:
Db 82 LBSGLQRGTIVBOCCTSGLYOLENYC 109

RESULT 14
 US-10-328-813-2
 : Sequence 2, Application US/10328813
 : Publication No. US20030113305A1
 : GENERAL INFORMATION:
 : APPLICANT: Osborne, William R.A.
 : APPLICANT: Ramesh, Nagarajan
 : TITLE OF INVENTION: Compositions and Methods for Treating Diabetes
 : PILE REFERENCE: P-UW 3264
 : CURRENT APPLICATION NUMBER: US/10/328,813
 : CURRENT FILING DATE: 2002-12-23
 : PRIOR APPLICATION NUMBER: US/09/185,852
 : PRIOR FILING DATE: 1998-11-04
 : PRIOR APPLICATION NUMBER: 60/087,660
 : PRIOR FILING DATE: 1998-06-02
 : NUMBER OF SEQ ID NOS: 11
 : SOFTWARE: PatentIn Ver. 2.0
 : SEQ ID NO 2
 : LENGTH: 110
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 : US-10-328-813-2

ORGANISM: Homo sapiens
US-10-328-813-2

Query Match	88.3%	Score 454;	DB 14;	Length 110;
Best Local Similarity	96.6%	Pred. No. 2.3e-42;		
Matches 85; Conservative	1;	Mismatches 2;	Indels 0;	Gaps 0;

Qy 8 SARFNPQHLGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLVGGSPGAGSLQPLA 67
:|||||
Db 22 AAAPNPQHLGSHLVEALYLVCGERGFFYPTKTRREADDPQGVQLVGGSPGAGSLQPLA 81

QY 68 LEGSLQKRGIVBQCCTSI CSLYLQLENYC 95
|||
Db 82 LEGLSLQKRGIVBQCCTSI CSLYLQLENYC 109

RESULT 15
US-10-346-563-2

Publication NO. US20030220229A1
GENERAL INFORMATION:
APPLICANT: Hickey, William F.
Griffin, Ann C.
TITLE OF INVENTION: Proinsulin peptide Compounds for Detecting and
Treating Type I Diabetes

NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, suite 51
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/346,563
FILING DATE: 16-Jan-2003
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/472,704
FILING DATE: 06-Jun-1995
APPLICATION NUMBER: US 08/272,220
FILING DATE: 08-JULY-1994
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: DCI-092
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 110 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-346-563-2

Query Match 88.3%; Score 454; DB 15; Length 110;
Best Local Similarity 96.6%; Pred. No. 2.3e-42;
Matches 85; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY	8	SARFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSIQPLA	67
	:		
DB	22	AAAFVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSIQPLA	81
	:		
QY	68	LEGSLOKRGIVEOCTSIQSYOLENYC	95
	:		
DB	82	LEGSLOKRGIVEOCTSIQSYOLENYC	109
	:		

Search completed: May 20, 2005, 00:44:43
Job time : 137 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using SW model

Run on: May 20, 2005, 00:32:40 ; Search time 38 Seconds
(without alignments)
243.074 Million cell updates/sec

Title: US-10-632-414-3

Perfect score: 514
Sequence: 1 ATSTGNSARFVNHGCSH.....IVEQCTSGSLYQLENYCG 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	460	89.5	110	2	B42179 insulin precursor
2	460	89.5	110	2	J00178 insulin precursor
3	454	88.3	110	1	IPHU insulin precursor
4	451	87.7	110	2	A42179 insulin precursor
5	411	80.0	110	1	INRB insulin precursor
6	404	78.6	110	1	IPDG insulin precursor
7	398	77.4	86	1	IPHO insulin precursor
8	398	77.4	110	1	INMS2 insulin 2 precursor
9	398	77.4	110	1	IPRT2 insulin 2 precursor
10	396	77.0	110	2	I48166 insulin precursor
11	389	75.7	110	1	IPRT1 insulin 1 precursor
12	387	75.3	84	1	IPPG insulin precursor
13	379	73.7	108	2	A39883 insulin precursor
14	370.5	72.1	105	1	IPBO insulin precursor
15	370	72.0	108	1	INMS1 insulin 1 precursor
16	338.5	65.9	108	2	S09278 insulin precursor
17	324.5	63.1	77	1	INSH insulin precursor
18	320	62.3	110	1	IPGP insulin precursor
19	287.5	55.9	103	2	I51221 insulin precursor
20	274.5	53.4	107	1	IPCH insulin precursor
21	271.5	52.8	106	1	IPXL2 insulin II precursor
22	268.5	52.2	106	1	IPXL1 insulin I precursor
23	265	51.6	109	1	IPRTDU insulin precursor
24	259.5	50.5	81	1	IPDK insulin precursor
25	252	49.0	96	2	PC7082 insulin precursor
26	250.5	48.7	51	1	INEL insulin - elephant
27	250.5	48.7	51	1	INWHF insulin - finback
28	250.5	48.7	51	1	INWHP insulin - sperm wh
29	248.5	48.3	51	1	INHY insulin - hamster

30	245.5	47.8	51	1	INMSGP insulin - Egyptian
31	244.5	47.6	51	2	A59151 insulin precursor
32	240.5	46.8	51	1	INCM4 insulin - Arabian
33	240.5	46.8	51	1	INGT insulin - goat
34	240.5	46.8	51	1	INWHIS insulin - sei whal
35	239.5	46.6	51	1	INCT insulin - cat
36	238.5	46.4	51	1	INMSQ insulin - common g
37	233.5	45.4	51	2	J00362 insulin - North Am
38	228.5	44.5	51	1	INCB insulin - Chinchi
39	225.5	43.9	51	1	INGG insulin - goose
40	221.5	43.1	51	1	INOS insulin - ostrich
41	221.5	43.1	51	1	INTK insulin - turkey (
42	221.5	43.1	51	1	A61129 insulin - black-be
43	221.5	43.1	51	1	INPO insulin - crested
44	221.5	43.1	51	2	A60414 insulin - slider t
45	219	42.6	52	2	S44470 insulin 12 - North

ALIGNMENTS

RESULT 1
B42179
insulin precursor - green monkey
C:Species: Cercopithecus aethiops (green monkey, grivet)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: B42179; A05232; S16494; S22056
R:Seino, S.; Bell, G.I.; Li, W.H.
Mol. Biol. Evol. 9, 193-203, 1992
A:Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
A:Reference number: A42179; MUID:92219553; PMID:1560757
A:Accession: B42179
A:Molecule type: DNA
A:Residues: 1-110 <SEI>
A:Cross-references: UNIPROT:P30407; EMBL:X61092; NID:G22808; PIDN:CAA43405.1; PID:G22809
A>Note: sequence extracted from NCBI backbone (NCBIN:95185, NCBI:P95194)
R:Peterson, J.D.; Nefflich, S.; Oyer, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A:Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsul
A:Reference number: A92111; MUID:72258016; PMID:4626369
A:Accession: A05232
A:Molecule type: protein
A:Residues: 57-87 <PRT>
C:Genetics:
C:Introns: 63/1
C:Superfamily: insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: insulin chain B #status predicted <BCH>
F:25-54,90-110/Product: insulin #status predicted <MAT>
F:57-87/Domain: connecting peptide #status experimental <CEP>
F:90-110/Domain: insulin chain A #status predicted <ACH>
F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 89.5%; Score 460; DB 2; Length 110;
Best Local Similarity 100.0%; Pred. No. 6,9e+43;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 FVNQHLGSHLYEALVYVCGRGFFYTPKTRREADPOVGQVELEGFGAGSLQPLALEG 70
DB 25 FVNQHLGSHLYEALVYVCGRGFFYTPKTRREADPOVGQVELEGFGAGSLQPLALEG 84

QY 71 SLQRRGIVEQCTSGSLYQLENYC 95
DB 85 SLQRRGIVEQCTSGSLYQLENYC 109

RESULT 2
J00178
insulin precursor - crab-eating macaque
C:Species: Macaca fascicularis (crab-eating macaque)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: J00178

R:Wetkäm, W.; Gronenberg, J.; Leineweber, M.; Wengenmayer, F.; Winnacker, E.L.
 Gene 19, 179-183, 1982
 A:Title: The nucleotide sequence of cDNA coding for preproinsulin from the primate Macaca
 A:Reference number: J00178; MUID:83080474; PMID:6184262
 A:Accession: J00178
 A:Molecule type: mRNA
 A:Residues: 1-110 <MBT>
 A:Cross-references: UNIPROT:P30406; GB:J00336; NID:9342121; PIDN:AAA56849.1; PID:9342122
 C:Superfamily: Insulin
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54,90-110/Product: insulin chain B #status predicted <MBT>
 F:55-54/Domain: insulin chain B #status predicted <MBT>
 F:55-89/Domain: insulin connecting C peptide #status predicted <CPT>
 F:90-110/Domain: insulin chain A #status predicted <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 89.5%; Score 460; DB 2; Length 110;
 Best Local Similarity 100.0%; Pred. No. 6,9e-43;
 Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 FVNQHLGSHLVEALYLVCGERGFFYPTKTRRAEDPQGVQLGGPGAGSLQPLALEG 70
 DB 25 FVNQHLGSHLVEALYLVCGERGFFYPTKTRRAEDPQGVQLGGPGAGSLQPLALEG 84

QY 71 SLQKRGIVQCCSTICSLYQLENYC 95
 DB 85 SLQKRGIVQCCSTICSLYQLENYC 109

RESULT 3
 IPIU
 Insulin precursor (validated) - human
 N:Alternate names: preproinsulin
 C:Species: Homo sapiens (man)
 C:Date: 23-Oct-1981 #sequence revision 23-Oct-1981 #text change 09-Jul-2004
 C:Accession: A93222; A94253; A93144; A92075; A91186; I58114; A01579; S58
 R:Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.
 Nature 288, 26-32, 1980
 A:Title: Sequence of the human insulin gene.
 A:Reference number: A93222; MUID:80120725; PMID:6243748
 A:Accession: A93222
 A:Molecule type: DNA
 A:Residues: 1-110 <BEL>
 A:Cross-references: UNIPROT:P01308; GB:J00265; NID:9186429; PIDN:AAA59172.1; PID:9386828
 R:Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.
 Science 209, 612-615, 1980
 A:Title: Genetic variation in the human insulin gene.
 A:Reference number: A94253; MUID:80236313; PMID:6248962
 A:Accession: A94253
 A:Molecule type: DNA
 A:Residues: 1-110 <UTL>
 A:Cross-references: GB:J00265; NID:9186429; PIDN:AAA59172.1; PID:9386828
 R:Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.
 Nature 282, 525-527, 1979
 A:Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.
 A:Reference number: A93216; MUID:80054779; PMID:503234
 A:Accession: A93216
 A:Molecule type: mRNA
 A:Residues: 1-110 <BEL>
 A:Cross-references: GB:J00265; NID:9186429; PIDN:AAA59172.1; PID:9386828
 R:Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.
 Science 208, 57-59, 1980
 A:Title: Nucleotide sequence of human preproinsulin complementary DNA.
 A:Reference number: A94251; MUID:80147417; PMID:6527840
 A:Accession: A94251
 A:Molecule type: mRNA
 A:Residues: 1-110 <SUA>
 A:Cross-references: GB:J00265; NID:9186429; PIDN:AAA59172.1; PID:9386828
 R:Nicol, D.S.H.W.; Smith, L.F.
 Nature 187, 483-485, 1960
 A:Title: Amino acid sequence of human insulin.
 A:Reference number: A93144
 A:Accession: A93144

A:Molecule type: protein
 A:Residues: 25-54;90-110 <NIC>
 R:Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.
 J. Biol. Chem. 246, 1375-1386, 1971
 A:Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pan
 A:Reference number: A92075; MUID:71116410; PMID:5101771
 A:Accession: A92075
 A:Molecule type: protein
 A:Residues: 57-87 <OYE>
 R:Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A:Title: Amino acid sequence of the C-peptide of human proinsulin.
 A:Reference number: A91186; MUID:71257722; PMID:5560404
 A:Accession: A91186
 A:Molecule type: protein
 A:Residues: 57-87 <KOA>
 R:Lucassen, A.M.; Juller, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell
 Nature Genet. 4, 305-310, 1993
 A:Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment
 A:Reference number: I58114; MUID:93364428; PMID:8358440
 A:Accession: I58114
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-59,63-110 <RES>
 A:Cross-references: GB:I5440; NID:9307071; PIDN:AAA59179.1; PID:9307072
 R:Sieber, P.; Kammer, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Ritcel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A:Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
 A:Reference number: A91636; MUID:7507277; PMID:4443293
 A:Contents: annotation; synthesis
 A:Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identi
 A:Note: article in German with English abstract
 R:Naltiani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A:Title: The synthesis of C-peptide of human proinsulin.
 A:Reference number: A91658; MUID:75040007; PMID:4803504
 A:Contents: annotation; synthesis of residues 57-87
 R:Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A:Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9
 A:Reference number: A90914
 A:Contents: annotation; synthesis of residues 57-87
 R:Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A:Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide juncti
 A:Reference number: S58661; MUID:96013185; PMID:7575420
 A:Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C:Genetics:
 A:Gene: GDB:INS
 A:Cross-references: GDB:119349; OMIM:176730
 A:Map position: 11p15.5-11p15.5
 A:Introns: 63/1
 C:Superfamily: Insulin
 C:Keywords: hormone; pancreas
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-54/Domain: insulin chain B #status experimental <MBT>
 F:57-87/Domain: insulin chain B #status experimental <MBT>
 F:57-87/Domain: connecting C peptide #status experimental <CPEP>
 F:90-110/Domain: insulin chain A #status experimental <ACH>
 F:31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 88.3%; Score 454; DB 1; Length 110;
 Best Local Similarity 96.6%; Pred. No. 3.1e-42;
 Matches 85; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 8 SARFVNHLCGSHLVEALYLVCGERGFFYPTKTRRAEDPQGVQLGGPGAGSLQPLA 67
 DB 22 AARFVNHLCGSHLVEALYLVCGERGFFYPTKTRRAEDPQGVQLGGPGAGSLQPLA 81

QY 68 LEGSLQKRGIVQCCSTICSLYQLENYC 95
 DB 82 LEGSLQKRGIVQCCSTICSLYQLENYC 109

RESULT 4
A42179
Insulin precursor - chimpanzee
C/Species: Pan troglodytes (chimpanzee)
C/Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C/Accession: A42179; S22058
R/Seino, S.; Bell, G.I.; Li, W.H.
Mol. Biol. Evol. 9, 193-203, 1992
A/Title: Sequences of primate insulin genes support the hypothesis of a slower rate of m
A/Reference numbers: A42179; MUID:92219553; PMID:1560757
A/Accession: A42179
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-110 <SEI>
A/Cross-references: UNIPROT:P30410; EMBL:X61089; NID:G38251; PIDN:CAA43403.1; PID:G38252
A/Note: Sequence extracted from NCBI backbone (NCBIF:95067)
C/Genetics:
A/Introns: 63/1
C/Superfamily: Insulin

Query Match 87.7%; Score 451; DB 2; Length 110;
Best Local Similarity 95.5%; Pred. No. 6.6e-42;
Matches 84; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 8 SARFVNQHLGSHLVKLVYVCGRGFFYTPKTRREADPQVQVGLGGPGAGSLQPLA 67
Db 22 ASARFVNQHLGSHLVKLVYVCGRGFFYTPKTRREADLQVQVGLGGPGAGSLQPLA 81
Qy 68 LEGSLQKRGIVEOCTSGSLYOLENYC 95
Db 82 LEGSLQKRGIVEOCTSGSLYOLENYC 109

RESULT 5
INRB
Insulin precursor - rabbit
N/Alternate names: preproinsulin
C/Species: Oryctolagus cuniculus (domestic rabbit)
C/Date: 24-Apr-1994 #sequence_revision 23-Aug-1997 #text_change 09-Jul-2004
C/Accession: A53438; A01581
R/Davekhar, S.U.; Giddings, S.J.; Rajkumar, P.A.; Carnaghi, L.R.; Menon, R.K.; Zahn, D.
J. Biol. Chem. 269, 8445-8454, 1994
A/Title: Insulin gene expression and insulin synthesis in mammalian neuronal cells.
A/Reference number: A53438; MUID:94179230; PMID:8132571
A/Accession: A53438
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-110 <DEV>
A/Cross-references: UNIPROT:P01311; GB:U03610; NID:G467970; PIDN:AAA19033.1; PID:G467971
R/Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A/Title: Species variation in the amino acid sequence of insulin.
A/Reference number: A90029; MUID:66160119; PMID:5949593
A/Accession: A01581
A/Molecule type: protein
A/Residues: 25-54; 90-110 <SMI>
C/Superfamily: Insulin
C/Key words: hormone; pancreas
F/1-24/Domain: insulin chain B #status predicted <SIG>
F/25-54/Domain: insulin chain B #status experimental <BCH>
F/25-54; 90-110/Product: insulin #status experimental <MAT>
F/57-87/Domain: connecting C peptide #status predicted <CEP>
F/90-110/Domain: insulin chain A #status experimental <ACH>
F/31-96; 43-109; 95-100/Disulfide bonds: #status predicted

Query Match 80.0%; Score 411; DB 1; Length 110;
Best Local Similarity 89.4%; Pred. No. 1.5e-37;
Matches 76; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 11 FVNQHLGSHLVKLVYVCGRGFFYTPKTRREADPQVQVGLGGPGAGSLQPLA 70
Db 25 FVNQHLGSHLVKLVYVCGRGFFYTPKTRREADLQVQVGLGGPGAGSLQPLA 84

Qy 71 SLOKRGIVEOCTSGSLYOLENYC 95
Db 85 ALOKRGIVEOCTSGSLYOLENYC 109

RESULT 6
IPDG
Insulin precursor - dog
C/Species: Canis lupus familiaris (dog)
C/Date: 24-Apr-1984 #sequence_revision 15-Nov-1984 #text_change 09-Jul-2004
C/Accession: A92413; A01587; S16493
R/Kwok, S.C.M.; Chan, S.J.; Steiner, D.F.
J. Biol. Chem. 258, 2357-2363, 1983
A/Title: Cloning and nucleotide sequence analysis of the dog insulin gene. Coded amino a
A/Reference number: A92413; MUID:83109071; PMID:6296142
A/Accession: A92413
A/Molecule type: DNA
A/Residues: 1-110 <SMI>
A/Cross-references: UNIPROT:P01321; GB:U00179; GB:U00042; NID:G994; PIDN:CAA23475.1; PID
R/Smith, L.F.
Am. J. Med. 40, 662-666, 1966
A/Title: Species variation in the amino acid sequence of insulin.
A/Reference number: A90029; MUID:66160119; PMID:5949593
A/Accession: A01587
A/Molecule type: protein
A/Residues: 25-54; 90-110 <SMIT>
R/Peterson, J.D.; Nehrlich, S.; Oyer, P.E.; Steiner, D.F.
J. Biol. Chem. 247, 4866-4871, 1972
A/Title: Determination of the amino acid sequence of the monkey, sheep, and dog proinsul
A/Reference number: A92111; MUID:72258016; PMID:4626369
A/Accession: S16493
A/Molecule type: protein
A/Residues: 65-85; 11, 87 <PET>
C/Superfamily: Insulin
C/Key words: hormone; pancreas
F/1-24/Domain: signal sequence #status predicted <SIG>
F/25-54/Domain: insulin chain B #status experimental <BCH>
F/25-54; 90-110/Product: insulin #status experimental <MAT>
F/57-87/Domain: connecting peptide #status predicted <CEP>
F/90-110/Domain: insulin chain A #status experimental <ACH>
F/31-96; 43-109; 95-100/Disulfide bonds: #status experimental

Query Match 78.6%; Score 404; DB 1; Length 110;
Best Local Similarity 88.2%; Pred. No. 8.5e-37;
Matches 75; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Qy 11 FVNQHLGSHLVKLVYVCGRGFFYTPKTRREADPQVQVGLGGPGAGSLQPLA 70
Db 25 FVNQHLGSHLVKLVYVCGRGFFYTPKTRREADLQVQVGLGGPGAGSLQPLA 84

Qy 71 SLOKRGIVEOCTSGSLYOLENYC 95
Db 85 ALOKRGIVEOCTSGSLYOLENYC 109

RESULT 7
IPHO
Insulin precursor - horse
C/Species: Equus caballus (domestic horse)
C/Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 09-Jul-2004
C/Accession: A01580; A92120
R/Harris, J.I.; Sanger, F.; Naughton, M.A.
Arch. Biochem. Biophys. 65, 427-428, 1956
A/Title: Species differences in insulin.
A/Reference number: A90082
A/Accession: A01580
A/Molecule type: protein
A/Residues: 1-30; 66-86 <HAR>
A/Cross-references: UNIPROT:P01310
R/Tager, H.S.; Steiner, D.F.
J. Biol. Chem. 247, 7936-7940, 1972
A/Title: Primary structures of the proinsulin connecting peptides of the rat and horse.


```

RESULT 10
148166      Insulin precursor - golden hamster
C:Species: Mesocricetus auratus (golden hamster)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Jul-1999
C:Accession: 148166
R:Belk, G.L.; Sanchez-Pescador, R.
Diabetes 33, 297-300, 1984
A>Title: Sequence of a cDNA encoding Syrian hamster preproinsulin.
A:Reference number: 148166; PMID:84133036; PMID:5365663
A:Accession: 148166
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M26328; NID:G191420; PIDN:AAA37089.1; PID:G305360
C:Superfamily: Insulin

Query Match      77.0%; Score 396; DB 2; Length 110;
Best Local Similarity 85.9%; Pred. No. 6.3e-36;
Matches 73; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY      11 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSIQPLALEG 70
DB      25 FVNQHLGSHLVEALYLVCGERGFFYTPKSRREVDPQVQLQLGCGGPAADLQTLALEV 84
QY      71 SLQKRGIVEQCTSIQSLYQLENYC 95
DB      85 AQQRKGIVDQCTSIQSLYQLENYC 109

RESULT 11
1PR11      Insulin 1 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 09-Jul-2004
C:Accession: A80788; A90789; A94231; B82120; I51945; A01589
R:Cordeiro, B.; Bell, G.; Tischer, E.; Denoto, F.M.; Ullrich, A.; Pictet, R.; Rutter, W.J.
Cell 18, 533-543, 1979
A>Title: Isolation and characterization of a cloned rat insulin gene.
A:Reference number: A80788; PMID:80045034; PMID:498283
A:Accession: A80788
A:Molecule type: DNA
A:Residues: 1-110 <CON>
A:Cross-references: UNIPROT:P01322; GB:J00747; NID:G204956; PIDN:AAA41442.1; PID:G204957
R:Lomedico, P.; Rosenthal, N.; Efstratiadis, A.; Gilbert, W.; Kolodner, R.; Tizard, R.
Cell 18, 545-558, 1979
A>Title: The structure and evolution of the two nonallelic rat preproinsulin genes.
A:Reference number: A80789; PMID:80045035; PMID:498284
A:Accession: A80789
A:Molecule type: DNA
A:Residues: 1-110 <LOW>
A:Cross-references: GB:J00747; NID:G204956; PIDN:AAA41442.1; PID:G204957
R:Steiner, D.F.; Clark, J.L.; Nolan, C.; Rubenstein, A.H.; Margolis, B.; Aten, B.; Oye
Recent Prog. Horm. Res. 25, 207-282, 1969
A>Title: Proinsulin and the biosynthesis of insulin.
A:Reference number: A94231; PMID:70067613; PMID:4311938
A:Accession: A94231
A:Molecule type: protein
A:Residues: 25-54; 90-110 <SNE>
R:Tager, H.S.; Steiner, D.F.
J. Biol. Chem. 247, 7936-7940, 1972
A>Title: Primary structures of the proinsulin connecting peptides of the rat and horse.
A:Reference number: A92120; PMID:73061498; PMID:4640931
A:Accession: B92120
A:Molecule type: protein
R:Lomedico, P.T.; Rosenthal, N.; Kolodner, R.; Efstratiadis, A.; Gilbert, W.
Ann. N.Y. Acad. Sci. 343, 425-432, 1980
A>Title: The structure of rat preproinsulin genes.
A:Reference number: I51945; PMID:80240379; PMID:6249167
A:Accession: I51945

```

```

A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-110 <RES>
A:Cross-references: GB:M2584; NID:G204947; PIDN:AAA41439.1; PID:G204948
C:Genetics:
A:Gene: INS1
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-54/Domain: Insulin chain B #status experimental <BCH>
F:57-87/Domain: Insulin #status experimental <MAT>
F:90-110/Domain: connecting peptide #status experimental <CRP>
F:31-96; 43-109; 95-100/Disulfide bonds: #status experimental

Query Match      75.7%; Score 389; DB 1; Length 110;
Best Local Similarity 84.7%; Pred. No. 3.6e-35;
Matches 72; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY      11 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSIQPLALEG 70
DB      25 FVNQHLGSHLVEALYLVCGERGFFYTPKSRREVDPQVQLQLGCGGPAADLQTLALEV 84
QY      71 SLQKRGIVEQCTSIQSLYQLENYC 95
DB      85 AQQRKGIVDQCTSIQSLYQLENYC 109

RESULT 12
1PR12      Insulin precursor - pig
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 22-Jun-1981 #sequence_revision 22-Jun-1981 #text_change 16-Jul-1999
C:Accession: A01583; A94572; S16492; A60835; B60835
R:Chance, R.E.; Ellis, R.M.; Bromer, W.W.
Biochem. 161, 165-167, 1968
A>Title: Porcine proinsulin: characterization and amino acid sequence.
A:Reference number: A94240; PMID:68286485; PMID:15657063
A:Accession: A01583
A:Molecule type: protein
A:Residues: 1-34; 'Q', 36-84 <CHA>
R:Chance, R.E.
submitted to the Atlas, July 1970
A:Reference number: A94572
A:Accession: A94572
A:Molecule type: protein
A:Residues: 1-84 <CH2>
R:Brown, H.; Sanger, F.; Kitai, R.
Biochem. J. 60, 565-565, 1955
A>Title: The structure of pig and sheep insulins.
A:Reference number: A90344
A:Accession: S16492
A:Molecule type: protein
A:Residues: 1-30; 31-51 <BRO>
R:Snell, L.; Damgaard, U.
Horm. Metab. Res. 20, 476-480, 1988
A>Title: Proinsulin heterogeneity in pigs.
A:Reference number: A60835; PMID:99032178; PMID:3181865
A:Accession: A60835
A:Molecule type: protein
A:Residues: 33-38; 40-62 <SNE>
A>Note: the authors report the characterization of a connecting peptide variant lacking 1
A:Accession: B60835
A:Molecule type: protein
A:Residues: 33-62 <SN2>
R:Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D.
Adv. Protein Chem. 26, 279-402, 1972
A>Title: Insulin: the structure in the crystal and its reflection in chemistry and biology
A:Reference number: A90017
C:Superfamily: Insulin
C:Keywords: hormone; pancreas
F:1-30/Domain: Insulin chain B #status experimental <BCH>

```



```

Insulin 1 precursor - mouse
C/Species: Mus musculus (house mouse)
C/Date: 24-Apr-1984 #sequence_revision 14-Jul-1994 #text_change 09-Jul-2004
C/Accession: B26342; A48172; A01592; B61012
R/Mentworth, B.M.; Schaefer, I.M.; Villa-Komaroff, L.; Chirgwin, J.M.
J. Mol. Evol. 23, 305-312, 1986
A/Title: Characterization of the two nonallelic genes encoding mouse preproinsulin.
A/Reference number: A92965; MUID:87169768; PMID:3104603
A/Accession: B26342
A/Molecule type: DNA
A/Residues: 1-108 <MEN>
A/Cross-references: UNIPROT:P01325; GB:X04725; NID:952712; PIDN:CA28434.1; PID:952713
R/Sawa, T.; Ohgaki, S.; Morioke, H.; Yano, S.
J. Mol. Endocrinol. 5, 61-67, 1990
A/Title: Molecular cloning and DNA sequence analysis of preproinsulin genes in the NON m
A/Reference number: A48172; MUID:90372989; PMID:2397023
A/Accession: A48172
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-108 <SAW>
R/Buenzli, H.F.; Glaetzer, B.; Kunz, P.; Muehlaupt, E.; Humbel, R.E.
Hoppe-Seyler's Z. Physiol. Chem. 353, 451-458, 1972
A/Title: Amino acid sequence of the two insulin from mouse (Mus musculus).
A/Reference number: A01592; MUID:72189455; PMID:5063718
A/Accession: A01592
A/Molecule type: protein
A/Residues: 25-54;88-108 <BUE>
R/Linde, S.; Nielsen, J.H.; Hansen, B.; Weidner, B.S.
J. Chromatogr. 462, 243-254, 1989
A/Title: Reversed-phase high-performance liquid chromatographic analyses of insulin bioe
A/Reference number: A61012; MUID:89292078; PMID:2661585
A/Accession: B61012
A/Molecule type: protein
A/Residues: 57-85 <LIN>
C/Superfamily: Insulin
C/Keywords: hormone; pancreas
F/1-24/Domain: signal sequence #status predicted <SIG>
F/25-54/Domain: insulin chain B #status experimental <BCH>
F/25-54;88-108/Product: Insulin #status experimental <MAT>
F/57-85/Domain: connecting peptide #status experimental <CEP>
F/88-108/Domain: insulin chain A #status experimental <ACH>
F/31-94;43-107;93-98/Disulfide bonds: #status predicted

Query Match 72.0%; Score 370; DB 1; Length 108;
Best Local Similarity 82.4%; Pred. No. 4.2e-33;
Matches 70; Conservative 4; Mismatches 9; Indels 2; Gaps 1;

QY 11 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLALEG 70
Db 25 FVKQHLGSPHLEALYLVCGERGFFYTPKSRREVEDPQVQLVGGSP--GDLQTLALEV 82
QY 71 SLQKRGIVGQCTSIQSLYQLENYC 95
Db 83 ARQKRGIVDQCTSIQSLYQLENYC 107

Search completed: May 20, 2005, 00:42:22
Job time : 39 secs

```

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 20, 2005, 00:23:45 ; Search time 175 Seconds
(without alignments)
280.912 Million cell updates/sec

Title: US-10-632-414-3
Performance: 514
Sequence: 1 ATTSTGNSARFVNHLGSH.....IVEQCTSIGSLYLENYCG 96

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	460	89.5	110	1	INS_CERAE
2	460	89.5	110	1	INS_MACPA
3	454	88.3	110	1	INS_GORGO
4	454	88.3	110	1	INS_HUMAN
5	451	87.7	110	1	INS_PANTR
6	450	87.5	110	1	INS_PONPY
7	411	80.0	110	1	INS_RABIT
8	408	79.4	110	1	INS_SPERT
9	404	78.6	110	1	INS_CANFA
10	398	77.4	86	1	INS_HORSE
11	398	77.4	110	1	INS2_MOUSE
12	398	77.4	110	1	INS2_MOUSE
13	396	77.0	110	1	INS2_MOUSE
14	389	75.7	110	1	INS2_MOUSE
15	387	75.3	108	1	INS2_MOUSE
16	381	74.1	110	1	INS2_MOUSE
17	379	73.7	108	1	INS2_MOUSE
18	375	73.0	110	2	INS2_MOUSE
19	370	72.0	105	1	INS2_MOUSE
20	370	72.0	108	1	INS2_MOUSE
21	366.5	71.3	105	1	INS2_MOUSE
22	338.5	65.9	108	1	INS2_MOUSE
23	335	65.2	65	2	INS2_MOUSE
24	335	65.2	65	2	INS2_MOUSE
25	320	62.3	110	1	INS2_MOUSE
26	287.5	55.9	103	1	INS2_MOUSE
27	274.5	53.4	107	1	INS2_MOUSE
28	271.5	52.8	106	1	INS2_MOUSE
29	268.5	52.2	106	1	INS2_MOUSE
30	265	51.6	109	1	INS2_MOUSE
31	259.5	50.5	81	1	INS2_MOUSE

32	252	49.0	96	2	Q7M0U6	Q7M0U6 bacillus br
33	250.5	48.7	51	1	INS_BALPH	P67973 balaenopter
34	250.5	48.7	51	1	INS_ELEMA	P01316 elephas max
35	250.5	48.7	51	1	INS_PHYCA	P67974 physeer ca
36	249.5	48.5	106	2	Q918Q7	Q918Q7 rana pipien
37	248.5	48.3	51	2	Q7M0G1	Q7M0G1 cricetidae
38	245.5	47.8	51	1	INS_ACOCA	P01324 acomya cali
39	244.5	47.6	51	2	Q7M217	Q7M217 canavalia e
40	240.5	46.8	51	1	INS_BALBO	P01314 balaenopter
41	240.5	46.8	51	1	INS_CAMPR	P01320 camelus dro
42	240.5	46.8	51	1	INS_CAMPR	P01319 capra hircu
43	239.5	46.6	51	1	INS_FELCA	P06306 felis silve
44	238.5	46.4	51	1	INS_SAISC	P67971 saimiri sci
45	233.5	45.4	51	1	INS_DIDWA	P18109 didelphis m

ALIGNMENTS

RESULT 1
INS_CERAE STANDARD; PRT; 110 AA.
ID INS_CERAE
AC P30407; P01309; (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DE 05-JUN-2004 (Rel. 44, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219953; PubMed=1560757;
RA Sano S., Bell G.I., Li W.;
RT "Sequences of primate insulin genes support the hypothesis of a slower
rate of molecular evolution in humans and apes than in monkeys.";
RL Mol. Biol. Evol. 9:193-203(1992).
RP [2]
RP SEQUENCE OF 57-87.
RX MEDLINE=7258016; PubMed=4626369;
RA Peterson J.D., Nehrlich S., Oyer P.E., Steiner D.F.;
RT "Determination of the amino acid sequence of the monkey, sheep, and
dog proinsulin C-peptides by a semi-micro Edman degradation
procedure.";
RL J. Biol. Chem. 247:4866-4871(1972).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
increases cell permeability to monosaccharides, amino acids and
fatty acids. It accelerates glycolysis, the pentose phosphate
cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>)
or send an email to license@ebi.ac.uk.
CC EMBL: X61092; CAA43405.1; -;
CC PIR: B42179; B42179;
DR HSSP; P01308; 1A10;
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1;
DR PRINTS; PR00277; INSULINB.
DR PRODOM; PD015667; MolIusc_ins; 1.
DR SMART; SMO0078; IIGF; 1.

```

DR PROSITE; PS00262; INSULIN; 1.
KW Direct protein sequencing; Glucose metabolism; Hormone;
KW Insulin family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12019 MW; 95A1F54BE7B247F9 CRC64;

Query Match 89.5%; Score 460; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.3e-39;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 FVNQHLGSHLYALVYVCGERGFFYTPKTRREADDPVQGVVLGGPGAGSLQPLALEG 70
DB 25 FVNQHLGSHLYALVYVCGERGFFYTPKTRREADDPVQGVVLGGPGAGSLQPLALEG 84
QY 71 SLQKRGIVEQCCTSIQSLYQLENYC 95
DB 85 SLQKRGIVEQCCTSIQSLYQLENYC 109

RESULT 2
INS_MACFA
ID INS_MACFA STANDARD; PRT; 110 AA.
AC P30706; P01309;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=63080474; PubMed=6184262; DOI=10.1016/0378-1119(82)90004-X;
RA Wetekam W., Gronenberg J., Leineweber M., Wengenmayer F.,
RA Minackner E.-L.;
RT "The nucleotide sequence of cDNA coding for preproinsulin from the
RT primate Macaca fascicularis."
RL Gene 19:179-183(1982).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00336; AAA36849.1; -.
DR PIR; J00178; J00178.
DR HSSP; P01308; IAI0.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULIN.
DR ProDom; PD015667; MolIusc_ins; 1.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.

```

```

KW Glucose metabolism; Hormone; Insulin family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;

Query Match 89.5%; Score 460; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 2.3e-39;
Matches 85; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 FVNQHLGSHLYALVYVCGERGFFYTPKTRREADDPVQGVVLGGPGAGSLQPLALEG 70
DB 25 FVNQHLGSHLYALVYVCGERGFFYTPKTRREADDPVQGVVLGGPGAGSLQPLALEG 84
QY 71 SLQKRGIVEQCCTSIQSLYQLENYC 95
DB 85 SLQKRGIVEQCCTSIQSLYQLENYC 109

RESULT 3
INS_GORGO
ID INS_GORGO STANDARD; PRT; 110 AA.
AC Q6YK33;
DT 25-OCT-2004 (Rel. 45, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Gorilla gorilla gorilla (Lowland gorilla).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Gorilla.
OX NCBI_TaxID=9595;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22833521; PubMed=12952878; DOI=10.1101/gr.948003;
RA Stead J.D.H., Hurler M.E., Jeffreys A.J.;
RL Genome Res. 13:2101-2111(2003).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY137500; AAN06935.1; -.
DR InterPro; IPR004825; Ins/IGF/relax.
DR ProDom; PD00049; Insulin; 1.
DR PRINTS; PR00277; INSULIN.
DR ProDom; PD015667; MolIusc_ins; 1.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Glucose metabolism; Hormone; Insulin family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin A chain.
FT DISULFID 31 96 Interchain (By similarity).

```


Query Match	88.3%	Score 454	DB 1	Length 110
Best Local Similarity	96.6%	Pred. No. 9.3e-39		
Matches	85	Conservative	1	Mismatches 2; Indels 0; Gaps 0;
QY	8	SARFVNHLTGGSHLVEALYLIVCGERGFYTPKTRRREDEPQVQVLELGGRGAGSLQPLA	67	
DB	22	AAAFVNHNLGGSHLVEALYLIVCGERGFYTPKTRRREDEPQVQVLELGGRGAGSLQPLA	81	
QY	68	LEGSLOKRGIVEOCCTGICSLYOLENCC	95	
DB	82	LEGSLOKRGIVEOCCTGICSLYOLENCC	109	
RESULT 5				
INS_PANTR				
INS_PANTR	STANDARD;	PRT;	110	AA.
AC	P30410;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-APR-1993 (Rel. 25, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	Insulin precursor.			
GN	Name=INS;			
OS	pan troglodytes (Chimpanzee).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.			
OX	NCB1:taxid=9598;			
LN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=9221953; PubMed=1560757;			
RA	Seino S., Bell G.I., Li W.;			
RT	"Sequence of primate insulin genes support the hypothesis of a slower			
RT	rate of molecular evolution in humans and apes than in monkeys."			
RL	Mol. Biol. Evol. 9:193-203 (1992).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=22831521; PubMed=12952878; DOI=10.1101/gr.948003;			
RA	Stead J.D.H., Hurler M.E., Jeffreys A.J.;			
RT	"Global haplotype diversity in the human insulin gene region."			
RL	Genome Res. 13:2101-2111(2003).			
CC	-1- FUNCTION: Insulin decreases blood glucose concentration. It			
CC	increases cell permeability to monosaccharides, amino acids and			
CC	fatty acids. It accelerates glycolysis, the pentose phosphate			
CC	cycle, and glycogen synthesis in liver.			
CC	-1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two			
CC	disulfide bonds.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- SIMILARITY: Belongs to the insulin family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X61089; CAA43403.1; -.			
DR	EMBL; AY137497; AAM06933.1; -.			
DR	PIR; A42179; A42179.			
DR	HSSP; P01308; 1A10.			
DR	InterPro; IPR004825; Ins/IGF/relax.			
DR	Pfam; PF00049; Insulin; 1.			
DR	PRINTS; PR00277; INSULINB.			
DR	ProDom; PD015667; Molusc_ins; 1.			
DR	PROSITE; PS00262; INSULIN_1.			
FM	Glucose metabolism; Hormone; Insulin family; Signal.			
FT	SIGNAL	1	24	
FT	CHAIN	25	54	Insulin B chain.
FT	PROPEP	57	87	C peptide.
FT	CHAIN	90	110	Insulin A chain.
FT	DISULFD	31	96	Interchain (By similarity).
FT	DISULFD	43	109	Interchain (By similarity).

FT DISULFID 95 100 By similarity.
 SQ SEQUENCE 110 AA; 12025 MM; 41EB8DF79837CEFS CRC64;
 Query Match 87.7%; Score 451; DB 1; Length 110;
 Best Local Similarity 95.5%; Pred. No. 1.9e-38;
 Matches 84; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 8 SARFVNHLCGSHLYEALVLCGERGFYPTKTRREADPQVGVEIGSGPGAGSLQPLA 67
 DB 22 ASAFVNHLCGSHLYEALVLCGERGFYPTKTRREADPQVGVEIGSGPGAGSLQPLA 81
 QY 68 LEGSLQKRGIVEQCTTSICSLYQLENYC 95
 DB 82 LEGSLQKRGIVEQCTTSICSLYQLENYC 109
 RESULT 6
 INS_PONPY STANDARD; PRT; 110 AA.
 AC OSHXV2;
 DT 05-JUL-2004 (Rel. 44, Created)
 DT 05-JUL-2004 (Rel. 44, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Pongo pygmaeus (Orangutan).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pongo.
 OC NCBI_TaxID=9600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22833521; PubMed=12952878; DOI=10.1101/gr.948003;
 RA Stead J.D.H., Hurles M.E., Jeffreys A.J.;
 RL "Global haplotype diversity in the human insulin gene region."
 CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the insulin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AY137503; AA06937.1; -
 DR HSBP; P01308; 1A10.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin.1.
 DR PRINTS; PR00277; INSULINB.
 DR PRODOM; PD015667; MolIusc_ins; 1.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Glucose metabolism; Hormone; Insulin family; Signal.
 FT SIGNAL 1 24 By similarity.
 FT CHAIN 25 54 Insulin B chain.
 FT PROPEP 57 87 C peptide.
 FT CHAIN 90 110 Insulin A chain.
 FT DISULFID 31 96 Interchain (By similarity).
 FT DISULFID 43 109 Interchain (By similarity).
 FT DISULFID 95 100 By similarity.
 SQ SEQUENCE 110 AA; 12038 MM; 22D2B32B94F520F8 CRC64;
 Query Match 87.5%; Score 450; DB 1; Length 110;
 Best Local Similarity 98.8%; Pred. No. 2.4e-38;
 Matches 84; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 FVNHLCGSHLYEALVLCGERGFYPTKTRREADPQVGVEIGSGPGAGSLQPLA 70
 DB 25 FVNHLCGSHLYEALVLCGERGFYPTKTRREADPQVGVEIGSGPGAGSLQPLA 84
 QY 71 SLQKRGIVEQCTTSICSLYQLENYC 95
 DB 85 SLQKRGIVEQCTTSICSLYQLENYC 109
 RESULT 7
 INS_RABIT STANDARD; PRT; 110 AA.
 AC P01311;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OC NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white; TISSUE=Pancreas;
 RX MEDLINE=94179230; PubMed=8132571;
 RA Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,
 RA Menon R.K., Zahm D.S.;
 RL "Insulin gene expression and insulin synthesis in mammalian neuronal
 RL cells."
 RT J. Biol. Chem. 269:8445-8454(1994).
 RN [2]
 RP SEQUENCE OF 25-54 AND 90-110
 RX MEDLINE=66160119; PubMed=5949593; DOI=10.1016/0002-9343(66)90145-8;
 RA Smith L.F.;
 RL "Species variation in the amino acid sequence of insulin."
 RL Am. J. Med. 40:662-666(1966).
 RN [3]
 RP SEQUENCE OF 56-110 FROM N.A.
 RA Giddings S.J., Carnaghi L.R., Devaskar S.U.;
 RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the insulin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U03610; AAA19033.1; -
 DR EMBL; M61153; AAA17540.1; -
 DR PIR; A53438; INRB.
 DR HSBP; P01308; 1B16.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin.1.
 DR PRINTS; PR00277; INSULINB.
 DR PRODOM; PD015667; MolIusc_ins; 1.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Direct protein sequencing; Glucose metabolism; Hormone;
 KW Insulin family; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.

FT PROPER 57 87 C peptide.
 FT CHAIN 90 110 Insulin A chain.
 FT DISULFID 31 96 Interchain.
 FT DISULFID 43 109 Interchain.
 FT DISULFID 95 100 Interchain.
 FT CONFLICT 83 83 E -> Y (in Ref. 3).
 SQ SEQUENCE 110 AA; 11838 MW; 82D2975B85D77FA8 CRC64;
 Query Match 80.0%; Score 411; DB 1; Length 110;
 Best Local Similarity 89.4%; Pred. No. 2.4e-34;
 Matches 76; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 11 FVNQHLGSHLYEALYLVGGERGFYTPKTRREADDPQVGVELGSGPGAGSLQPLALEG 70
 DB 25 FVNQHLGSHLYEALYLVGGERGFYTPKSRREVEELQVGQALGGGPGAGLQPEALLL 84
 QY 71 SLQKRGIVGQCCTSIQSLYQLENYC 95
 DB 85 ALQKRGIVGQCCTSIQSLYQLENYC 109
 RESULT 8
 INS_SPECTR STANDARD; PRT; 110 AA.
 ID INS_SPECTR
 AC Q91X13;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
 OC Spermophilus.
 NC NCBI_TaxID=43179;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Tissue=Pancreas;
 RA Tredrea M.M., Buck M.J., Gnanapavan J., Squire T.L., Andrews M.T.;
 RL "Regulation of PDK4 expression in a hibernating mammal.";
 RT Submitted (JUN-2001) to the EMBL/GenBank/DBD databases.
 CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AY038604; AAK72558.1; -;
 DR HSSP; P01308; IEV6.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PRO0277; INSULINB.
 DR ProDom; PD015667; Mollusc_ins; 1.
 DR SMART; SM00078; IIGF; 1.
 DR KMW Glucose metabolism; Hormone; Insulin family; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.
 FT PROPER 57 87 C peptide.
 FT CHAIN 90 110 Insulin A chain.
 FT DISULFID 31 96 Interchain (By similarity).
 FT DISULFID 43 109 Interchain (By similarity).

FT DISULFID 95 100 By similarity.
 SQ SEQUENCE 110 AA; 12004 MW; 4511768D622BEE5 CRC64;
 Query Match 79.4%; Score 408; DB 1; Length 110;
 Best Local Similarity 89.4%; Pred. No. 4.8e-34;
 Matches 76; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
 QY 11 FVNQHLGSHLYEALYLVGGERGFYTPKTRREADDPQVGVELGSGPGAGSLQPLALEG 70
 DB 25 FVNQHLGSHLYEALYLVGGERGFYTPKSRREVEELQVGQALGGGPGAGLQPEALLL 84
 QY 71 SLQKRGIVGQCCTSIQSLYQLENYC 95
 DB 85 ALQKRGIVGQCCTSIQSLYQLENYC 109
 RESULT 9
 INS_CANFA STANDARD; PRT; 110 AA.
 ID INS_CANFA
 AC P01321;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Insulin precursor.
 GN Name=INS;
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NC NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83109071; PubMed=6296142;
 RA Kwok S.C.M., Chan S.J., Steiner D.F.;
 RT "Cloning and nucleotide sequence analysis of the dog insulin gene.
 RT Coded amino acid sequence of canine preproinsulin predicts an
 RT additional C-peptide fragment.";
 RJ J. Biol. Chem. 258:2357-2363(1983).
 RN [2]
 RP SEQUENCE OF 25-54 AND 90-110.
 RX MEDLINE=66160119; PubMed=5949593; DOI=10.1016/0002-9343(66)90145-8;
 RA Smith U.F.;
 RT "Species variation in the amino acid sequence of insulin.";
 RL Am. J. Med. 40:662-666(1966).
 CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; V00179; CAA23475.1; -;
 DR PIR; A92413; IPDG.
 DR HSSP; P01317; IAPH.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PRO0277; INSULINB.
 DR ProDom; PD015667; Mollusc_ins; 1.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 DR KMW Direct protein sequencing; Glucose metabolism; Hormone; Insulin family; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.

FT	PROPEP	57	87	C peptide.
FT	CHAIN	90	110	Insulin A chain.
FT	DISULFID	31	96	Interchain.
FT	DISULFID	43	109	Interchain.
FT	DISULFID	95	100	
8Q	SEQUENCE	110 AA;	12190 MW;	A574791864A4FB98 CRC64;
Query Match				
	Best Local Similarity	78.6%;	Score 404;	DB 1;
	Matches 75;	Conservative 1;	Mismatches 9;	Indels 0;
OY				
Db	11 FVNOHLCSHLYEALYLVCGERGFPYTPKTRREAPDPVGVYELDGGSPGASLOPLALEG	70		
OY	25 FVNOHLCSHLYEALYLVCGERGFPYTPKTRREAPDPVGVYELDGGSPGASLOPLALEG	84		
Db	71 SLQKRGIVEQCCTSIQSLPYOLENYC	95		
	85 ALQKRGIVEQCCTSIQSLPYOLENYC	109		
RESULT 10				
INS_HORSE				
ID	INS_HORSE	STANDARD;	PRT;	86 AA.
AC	P01310;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	25-OCT-2004 (Rel. 45, Last annotation update)			
DE	Insulin precursor.			
GN	Name=INS;			
OS	Equus caballus (Horse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Perissodactyla; Equidae; Equus.			
OX	NCBI_TaxID=9796;			
RN	[1]			
RP	SEQUENCE OF 1-30 AND 66-86.			
RX	PubMed=13373434;			
RA	Harris J.I., Sanger F., Naughton M.A.;			
RL	"Species differences in insulin."			
RL	Arch. Biochem. Biophys. 65:427-438(1956).			
RP	[2]			
RP	SEQUENCE OF 33-63.			
RX	MEDLINE=73061498; PubMed=4640931;			
RA	Tager H.S., Steiner D.F.;			
RT	"Primary structures of the proinsulin connecting peptides of the rat			
RT	and the horse."			
RL	J. Biol. Chem. 247:7936-7940(1972).			
CC	-1- FUNCTION: Insulin decreases blood glucose concentration. It			
CC	increases cell permeability to monosaccharides, amino acids and			
CC	fatty acids. It accelerates glycolysis, the pentose phosphate			
CC	cycle, and glycogen synthesis in liver.			
CC	-1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two			
CC	disulfide bonds.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- SIMILARITY: Belongs to the insulin family.			
CC	-1- CAUTION: X's at positions 31-32 and 64-65 represent paired basic			
CC	residues assumed by homology to be present in the precursor			
CC	molecule.			
CC	PIR: A01580; IPHO.			
DR	HSGP; P01317; IAPH.			
DR	InterPro: IPR004825; Ins/IGF/relax.			
DR	Pfam; PF00049; Insulin; 1.			
DR	PRINTS; PR00277; INSULINB.			
DR	SMART; SM00078; IIGF; 1.			
DR	PROSITE; PS00262; INSULIN; 1.			
KW	Direct protein sequencing; Glucose metabolism; Hormone;			
KW	Insulin family.			
FT	CHAIN	1	30	Insulin B chain.
FT	PROPEP	33	63	C peptide.
FT	CHAIN	66	86	Insulin A chain.
FT	DISULFID	7	72	Interchain.
FT	DISULFID	19	85	Interchain.
FT	DISULFID	71	76	
8Q	SEQUENCE	86 AA;	9142 MW;	A3E1E822711BDB46 CRC64;

```

Query Match      77.4%; Score 398; DB 1; Length 86;
Best Local Similarity 85.9%; Pred. No. 4e-33;
Matches 73; Conservative 1; Mismatches 11; Indels 0; Gaps 0

Qy      11 FVNHGLGSHLVEALYLIVCGERGPFTYRDKTRREADPPQGVGVETLGGFGAGSLQPLALEG 70
Db      1 FVNHGLGSHLVEALYLIVCGERGFPTYRDKAXXXEADDPQGVGVETLGGFGAGSLQPLALG 60
Qy      71 SLQKRGIVGECCTSLCSLYOLENTC 95
Db      61 PQQXGIVECCCTGICSLYOLENTC 85

RESULT 11
INS2_MOUSE STANDARD; PRT; 110 AA.
ID INS2_MOUSE
AC P01326;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Insulin 2 precursor.
GN Name=Ins2; Synonym=Ins-2;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1]
RP SEQUENCE FROM N.A.
RP MEDLINE=87169768; PubMed=3104603;
RA Wertworth B.M., Schaefer I.M., Villa-Komaroff L., Chirgwin J.M.;
RT "Characterization of the two nonallelic genes encoding mouse
RL preproinsulin."
RL J. Mol. Evol. 23:305-312 (1986).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=NON;
RC MEDLINE=90372989; PubMed=2397023;
RX MEDLINE=90372989; PubMed=2397023;
RA Sawa T., Ohgaku S., Morioka H., Yano S.;
RT "Molecular cloning and DNA sequence analysis of preproinsulin genes in
RT the NON mouse, an animal model of human non-obese, non-insulin-
RL dependent diabetes mellitus."
RL J. Mol. Endocrinol. 5:61-67 (1990).
[3]
RP SEQUENCE OF 25-54 PND 90-110.
RP MEDLINE=72189455; PubMed=5063718;
RA Buzenli H.F., Glatthaar B., Kunz P., Muehlaupt E., Hummel R.E.;
RT "Amino acid sequence of the two insulins from mouse (Mus musculus).";
RL Hoppe-Seyler's Z. Physiol. Chem. 353:451-458 (1972).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
-----
CC EMBL; X04724; CAA28433.1; -.
CC PIR; A26342; INMS2.
CC HSSP; P01317; IAPB.
CC MGD; MGI:96573; Ins2.
CC GO; GO:0005634; C:nucleus; IDA.
CC GO; GO:0005732; C:nuclear nucleolar ribonucleoprotein complex; IDA.
CC GO; GO:0000187; P:activation of MAPK; IDA.

```

```

DR GO; GO:0006006; P:glucose metabolism; IMP.
DR GO; GO:0008286; P:insulin receptor signaling pathway; IDA.
DR GO; GO:0016042; P:lipid catabolism; IDA.
DR GO; GO:0042281; P:regulation of apoptosis; IMP.
DR GO; GO:0042325; P:regulation of phosphorylation; IDA.
DR GO; GO:0006983; P:response to ER-overload; IMP.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR ProDom; PD015667; Mollusc_ins; 1.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
DR Direct protein sequencing; Glucose metabolism; Hormone;
KW Insulin family; Multigene family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin 2 B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin 2 A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100 Interchain.
SQ SEQUENCE 110 AA; 12364 MW; 3554C8803D24FDAD CRC64;

Query Match 77.4%; Score 398; DB 1; Length 110;
Best Local Similarity 85.9%; Pred. No. 5,1e-33;
Matches 73; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 11 FVNQHLGSHLYVEALVYVCGRGFFPTKTRREADPPVQGVYGGGPGAGSLQPLALEG 70
DB 25 FVKQHLGSHLYVEALVYVCGRGFFPTKTRREADPPVQGVYGGGPGAGDLQTLALEV 84

QY 71 SLOKRGIVEQCCTSCISLYOLENYC 95
DB 85 AQKRGIVDQCCTSCISLYOLENYC 109

RESULT 12
INS2 RAT STANDARD; PRT; 110 AA.
ID INS2 RAT
AC P01333;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Insulin 2 precursor.
GN Name=Ins2; Synonyms=Ins-2;
OS Rattus norvegicus (Rat);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=80045035; PubMed=498284; DOI=10.1016/0092-8674(79)90071-0;
RA Lomedico P., Rosenblatt N., Efstratiadis A., Gilbert W., Kolodner R.,
RA Tizard R.;
RT "The structure and evolution of the two nonallelic rat preproinsulin
RT genes.";
RT Cell 18:545-558(1979).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86310882; PubMed=2427930;
RA Soares M.B., Schin E., Henderson A., Karathanasis S.K., Cate R.,
RA Zeitlin S., Chirgwin J., Efstratiadis A.;
RT "RNA-mediated gene duplication: the rat preproinsulin I gene is a
RT functional retroposon.";
RT Mol. Cell. Biol. 5:2090-2103(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80240379; PubMed=6249167;
RA Lomedico P.T., Rosenthal N., Kolodner R., Efstratiadis A., Gilbert W.,
RT "The structure of rat preproinsulin genes.";
RT Ann. N. Y. Acad. Sci. 343:425-432(1980).
RN [4]

```

```

RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=70067613; PubMed=4311938;
RA Steiner D.F., Clark J.L., Nolan C., Rubenstein A.H., Margoliash E.,
RA Aten B., Oyer P.E.;
RT "Protein and the biosynthesis of insulin.";
RT Recent Prog. Horm. Res. 25:207-282(1969).
RN [5]
RP SEQUENCE OF 57-87.
RX MEDLINE=73061498; PubMed=4640931;
RA Tager H.S., Steiner D.F.;
RT "Primary structures of the proinsulin connecting peptides of the rat
RT and the horse.";
RT J. Biol. Chem. 247:7936-7940(1972).
RN [6]
RP SEQUENCE OF 57-87, AND REVISIONS.
RX MEDLINE=72177385; PubMed=4554104;
RA Markussen J., Sundby F.;
RT "Rat-proinsulin C-peptides: Amino-acid sequences.";
RT Eur. J. Biochem. 25:153-162(1972).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SIMILARITY: Belongs to the insulin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; V01243; CAA24560.1; -.
DR EMBL; J00748; AAA41443.1; -.
DR EMBL; M25585; AAA41440.1; -.
DR EMBL; M25583; AAA41440.1; JOINED.
DR PIR; B90789; IPR2.
DR HSSP; P01317; IAPH.
DR RGD; 2916; Ins2.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR ProDom; PD015667; Mollusc_ins; 1.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Direct protein sequencing; Glucose metabolism; Hormone;
KW Insulin family; Multigene family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin 2 B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin 2 A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100 Interchain.
SQ SEQUENCE 110 AA; 12339 MW; 3A626DA98C86F3CA CRC64;

Query Match 77.4%; Score 398; DB 1; Length 110;
Best Local Similarity 85.9%; Pred. No. 5,1e-33;
Matches 73; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 11 FVNQHLGSHLYVEALVYVCGRGFFPTKTRREADPPVQGVYGGGPGAGSLQPLALEG 70
DB 25 FVKQHLGSHLYVEALVYVCGRGFFPTKTRREADPPVQGVYGGGPGAGDLQTLALEV 84

QY 71 SLOKRGIVEQCCTSCISLYOLENYC 95
DB 85 AQKRGIVDQCCTSCISLYOLENYC 109

```

```

RESULT 13
INS_CRILLO STANDARD; PRT; 110 AA.
ID INS_CRILLO
AC P01313:
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Crictellus longicaudatus (long-tailed hamster) (Chinese hamster);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Crictellus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84133036; PubMed=6365663;
RA Bell G.I., Sanchez-Pescador R.;
RT "Sequence of a cDNA encoding Syrian hamster preproinsulin.";
RL Diabetes 33:297-300(1984).
RN [2]
RP SEQUENCE OF 25-54 AND 90-110.
RA Neelon F.A., Delcher H.K., Steinman H., Lebovitz H.E.;
RT "Structure of hamster insulin: comparison with a tumor insulin.";
RL Fed. Proc. 32:300-300(1973).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M26328; AAA37089.1; .
DR HSBP; P01308; IEV6.
DR InterPro; IPR004825; Ins/IGF/relex.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR PRODom; PD015667; MolIuc_Ins; 1.
DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Direct protein sequencing; Glucose metabolism; Hormone;
KW Insulin family; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 Insulin B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100
SQ SEQUENCE 110 AA; 12268 MW; 219E92B85A353CEC CRC64;
Query Match 77.0%; Score 396; DB 1; Length 110;
Beef Local Similarity 85.9%; Pred.No. 8,2e-33;
Matches 73; Conservative 4; Mismatches 8; Indels 0; Gaps 0;
QY 11 FVNQHLCSHLYEALYLVGSGRPFYTPKTRREADPOVGVEYAGSGAGSLQPLALEG 70
DB FVNQHLCSHLYEALYLVGSGRPFYTPKTRREADPOVGVEYAGSGAGSLQPLALEV 84
QY 71 SLQKRGIVQCCSTICSLYOLENYC 95
DB SLQKRGIVQCCSTICSLYOLENYC 109

```

```

RESULT 14
INSI_RAT STANDARD; PRT; 110 AA.
ID INSI_RAT
AC P01322:
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Insulin 1 precursor.
GN Name=Ins1; Synonym=Ins-1;
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80045034; PubMed=498283; DOI=10.1016/0092-8674(79)90070-9;
RA Cordell B., Bell G.I., Tischer E., Denoto F.M., Ullrich A.,
RT "Isolation and characterization of a cloned rat insulin gene.";
RL Cell 18:533-543(1979).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Liver;
RX MEDLINE=80045035; PubMed=498284; DOI=10.1016/0092-8674(79)90071-0;
RA Pictet R.L., Rutter W.J., Goodman H.M.;
RT "Isolation and characterization of a cloned rat insulin gene.";
RL Cell 18:545-558(1979).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80240379; PubMed=6249167;
RA Lomedico P.T., Rosenthal N., Kolodner R., Efstratiadis A., Gilbert W.;
RT "The structure of rat preproinsulin genes.";
RN Ann. N.Y. Acad. Sci. 343:425-432(1980).
RN [4]
RP SEQUENCE OF 25-54 AND 90-110.
RX MEDLINE=70067613; PubMed=4311938;
RA Steiner D.F., Clark J.V., Nolan C., Rubenstein A.H., Margolis E.,
RA Aten B., Oyer P.E.;
RT "Proinsulin and the biosynthesis of insulin.";
RL Recent Prog. Horm. Res. 25:207-282(1969).
RN [5]
RP SEQUENCE OF 57-87.
RX MEDLINE=73061498; PubMed=4640931;
RA Tager H.S., Steiner D.F.;
RT "Primary structures of the proinsulin connecting peptides of the rat
RT and the horse.";
RL J. Biol. Chem. 247:7936-7940(1972).
RN [6]
RP SEQUENCE OF 57-87, AND REVISIONS.
RX MEDLINE=72177385; PubMed=4554104;
RT Markussen J., Sundby F.;
RT "Rat-proinsulin C-peptides. Amino-acid sequences.";
RL Eur. J. Biochem. 25:153-162(1972).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

DR EMBL; V01242; CAA24559.1; -
DR EMBL; J00747; AAA41442.1; -
DR EMBL; M25584; AAA41439.1; -
DR PIR; A90788; IPTL.
DR HSSP; P01308; 1A7F.
DR RGD; 2915; Ins1.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR ProDom; PD015667; Mollusc_ins; 1.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Direct protein sequencing; Glucose metabolism; Hormone;
KW Insulin family; Multigene family; Signal.
FT CHAIN 1 24 Insulin 1 B chain.
FT PROPEP 57 87 C peptide.
FT CHAIN 90 110 Insulin 1 A chain.
FT DISULFID 31 96 Interchain.
FT DISULFID 43 109 Interchain.
FT DISULFID 95 100 Interchain.
SQ SEQUENCE 110 AA; 12420 MW; 51D606DA54AE353 CRC64;

Query Match 75.7%; Score 389; DB 1; Length 110;
Best Local Similarity 84.7%; Pred. No. 4.3e-32;
Matches 72; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 11 FVNQHCGSHLVKLVKVGEGEPFPTKTRREAEDPOVGOVLGGPGAGSLQPLALBG 70
DB 25 FVQHCHGPHLVKLVKVGEGEPFPTKTRREAEDPOVGOVLGGPGAGSLQPLALBG 84
QY 71 SLQKRGIVBQCTCSISLYOLENYC 95
DB 85 ARQKRGIVDCCCTCSISLYOLENYC 109

RESULT 15
ID INS_PIG STANDARD; PRT; 108 AA.
AC P01315; O9TSJ5;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Insulin precursor.
GN Name=INS;
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1] _
RP SEQUENCE FROM N.A.
RA Han X.G., Tsch B.E.;
RT "Complete porcine preproinsulin cDNA sequence."
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RN [2] _
RP SEQUENCE FROM N.A.
RA STRAIN=Large white;
RX MEDLINE=22135958; PubMed=12140686; DOI=10.1007/s00035-001-3059-x;
RA Amarger V., Nguyen M., Laere A.S., Braunschweig M., Nezer C.,
RA Georges M., Anderson L.;
RT "Comparative sequence analysis of the INS-IGF2-H19 gene cluster in
RL pig.";
RN Mamm. Genome 13:388-398 (2002).
RP [3] _
RP SEQUENCE OF 25-108.
RX MEDLINE=68286485; PubMed=5657063;
RA Chance R.E., Ellis R.M., Brower W.W.;
RT "Porcine proinsulin: characterization and amino acid sequence."
RL Science 161:165-167 (1968).
RN [4] _
RP REVISION TO 59.
RA Chance R.E.;
RL Submitted (JUL-1970) to the PIR data bank.

RN [5] _
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;
RT "Insulin. The structure in the crystal and its reflection in chemistry
RL and biology".
RN Adv. Protein Chem. 26:279-402 (1972).
RN [6] _
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Isaacs N.W., Agarwal R.C.;
RT "Experience with fast Fourier least squares in the refinement of the
RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A
RT resolution".
RL Acta Crystallogr. A 34:782-791 (1978).
RN [7] _
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=8909318; PubMed=2805485;
RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,
RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,
RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;
RT "The structure of 22n pig insulin crystals at 1.5-A resolution."
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456 (1988).
RN [8] _
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=92126280; PubMed=1772633; DOI=10.1107/S010876819100842X;
RA Balchmidt P., Hansen F.B., Dodson E., Dodson G., Korfner F.;
RT "Structure of porcine insulin cocrystallized with clupeine Z."
RL Acta Crystallogr. B 47:975-986 (1991).
RN [9] _
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
RX MEDLINE=9122450; PubMed=2025410; DOI=10.1107/S0108768190009570;
RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J.,
RA Dodson G.G., North A.C.T.;
RT "Structure of the pig insulin dimer in the cubic crystal."
RL Acta Crystallogr. B 47:127-136 (1991).
RN [10] _
RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RX PubMed=15299880;
RA Diau J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT "Structure of monomeric porcine DesB1-B2 desptapeptide (B26-B30)
RT insulin at 1.65-A resolution".
RL Acta Crystallogr. D 53:507-512 (1997).
CC -1- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -1- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the insulin family.
CC -1- DATABASE: NAME=Protein Spotlight; NOTE=Issue 9 of April 2001;
CC WWW="http://www.expasy.org/spotlight/articles/sp1009.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ibt-sib.ch/announce/>
CC or send an email to license@ibt-sib.ch).
CC -----
DR EMBL; AF064555; AAC77920.1; ALT INIT.
DR EMBL; AY044828; AAL69550.1; -
DR PDB; 1DEI; X-ray; A/C=88-108, B/D=25-47.
DR PDB; 1I2A; X-ray; A/C=88-108, B/D=25-54.
DR PDB; 1I2B; X-ray; A/C=88-108, B/D=25-54.
DR PDB; 1WPU; X-ray; A/C=88-108, B/D=25-54.
DR PDB; 1WAV; X-ray; A=88-108, B=27-49.
DR PDB; 1WAV; X-ray; A/C=88-108, B/D=25-54.
DR PDB; 1ZNI; X-ray; A/B/C/D/E/F=25-108, B/D/F/H/J/L=25-54.
DR PDB; 1ZNI; X-ray; A/C=88-108, B/D=25-54.
DR PDB; 2TCI; X-ray; A/C=88-108, B/D=25-54.
DR PDB; 3INT; Neutron; A/C=88-108, B/D=25-54.
DR PDB; 3MTH; X-ray; A/C=88-108, B/D=25-54.

DR PDB; 4INS; X-ray; A/C=88-108, B/D=25-54.
 DR PDB; 6INS; X-ray; E/F=25-108.
 DR PDB; 7INS; X-ray; A/C/E=88-108, B/D/F=25-54.
 DR PDB; 9INS; X-ray; A=88-108, B=25-54.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR ProDom; PD015667; Mollusc_ins; 1.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW 3D-structure; Direct protein sequencing; Glucose metabolism; Hormone;
 KW Insulin family; Signal.
 KM SIGNAL 1 24
 FT CHAIN 25 54 Insulin B chain.
 FT PROPEP 57 85 C peptide.
 FT CHAIN 88 108 Insulin A chain.
 FT DISULFID 31 94 Interchain.
 FT DISULFID 43 107 Interchain.
 FT DISULFID 93 98
 FT HELIX 26 46
 FT STRAND 48 48
 FT STRAND 89 94
 FT HELIX 100 106
 FT STRAND 107 107
 SQ SEQUENCE 108 AA; 11671 MW; CB4491B429858EBE CRC64;

Query Match 75.3%; Score 387; DB 1; Length 108;
 Best Local Similarity 87.1%; Pred. No. 6.8e-32;
 Matches 74; Conservative 1; Mismatches 8; Indels 2; Gaps 1;

QY 11 FVNQHLCGSHLVEALYLVGGERGFFYTPKTRREAEDPQGVVELGGPAGSLOPLALEG 70
 DB 25 FVNQHLCGSHLVEALYLVGGERGFFYTPKARRAENPQAGAVELGG--GLGGLQALALEG 82

QY 71 SLQKRGIVQCCCTSGISLYQLENYC 95
 DB 83 PPQKRGIVQCCCTSGISLYQLENYC 107

Search completed: May 20, 2005, 00:38:50
 Job time : 176 secs

THIS PAGE LEFT BLANK